Curious Curves

Ellipses, hyperbolae and other geometric wonders

Everything is mathematical

Curious Curves

Curious Curves

Ellipses, hyperbolae and other geometric wonders

Josep Sales – Francesc Banyuls

Everything is mathematical

To my dear Marta, Dèlia, Otger, Bernat and Pau
JS

To the three girls of the house, Aina, Mar and Sònia
FB

Published by RBA Coleccionables, S.A.
c/o Hothouse Developments Ltd
91 Brick Lane, London, E1 6QL

Localisation: Windmill Books Ltd.

ISSN: 2050-649X

Printed in Spain

Contents

Preface

We have often heard it said that the shortest line between two points is a straight line. Curves are not only a longer route between two points but also have a distinct, even unique form compared to the conformity of straight lines. If we were to stop and look at our surroundings we would see that there are more curves in everyday life than straight lines.

From the dangerous bends of a mountain road to the multitude of spiral families, there is a great variety of curves with an immense number of applications. Curves are evidence of the immensity of thought, of mental systematisation and the capacity of the human mind. Curves are trajectories of great ideas, drawings, mathematical expressions and predictions of true-life phenomena. They are zigzagging routes of properties and relationships with specific parametric values. Simple curves appear when resolving difficult problems and families of curves appear as possible solutions in complicated situations. Curves have a rebellious and flexible spirit; they can be open or closed and can be reduced to the simplest curve of all – the straight line! Curves imply a breakthrough in mathematical thought, the gateway to an infinite universe of relationships, forms, surprises, families, relatives and representations.

Since ancient times, curves have attempted to be an explanation of dimensions both in human terms and with the universe of numbers. They define car wheels, trajectories of space rockets, probable movements of electrons in an atomic nucleus and the paths of hurricanes. They construct cathedral domes and arches, sculpture figures and constructive forms in modern museums.

Man attempts to control curves by forcing them to explain themselves by means of numbers, equations… and applying the most powerful mathematical knowledge to them, such as infinitesimal calculus, differential equations and elliptic integrals.

Curves are presented as volatile structures, evolving, involute, inverse, pedals, lemniscates, cycloids, concoids, spirals, strofoids, rolling, sliding, deltoids, astroids, roses… an infinite and marvellous ensemble.

The reader can approach this book without needing to follow the order of the chapters and can also save themselves from reading sections they may consider to be excessively mathematical without losing overall comprehension.

We would like to highlight the great help that using some of the magnificent new curve representation and symbolic calculation programs, such as Geogebra and

Derive, have provided to the development of this book. After reading it, we would recommend that the reader has as good a time with these programs as we did, and our ultimate intention is that the reader takes a peep through the keyhole to the house of curves to discover an immensity of marvels and surprising mathematical risks.

Chapter 1

What Are Curves Used For?

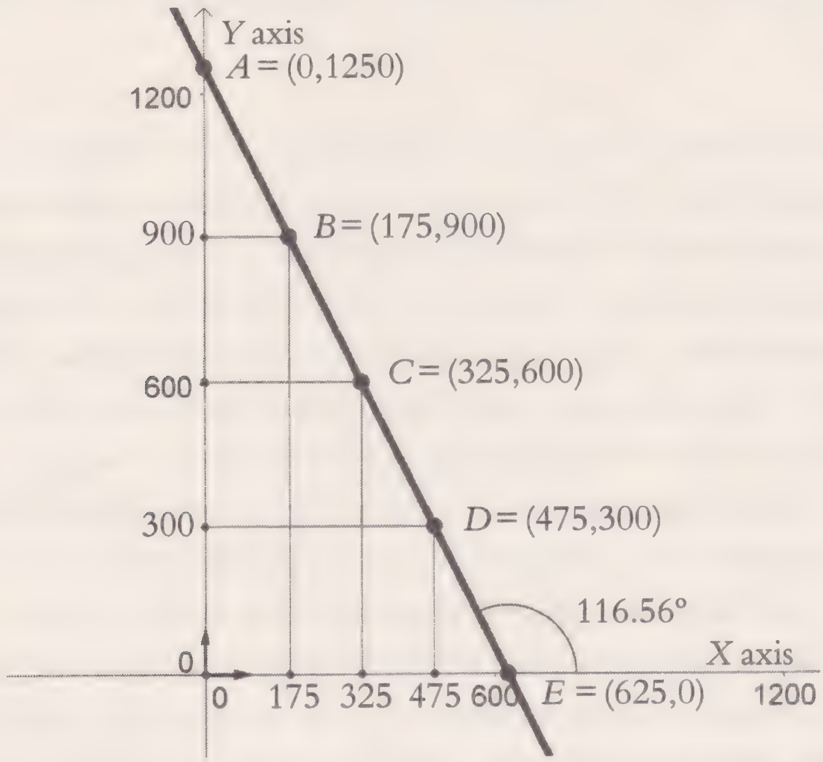
Curves are used in all fields of science and technology and in everyday life. This chapter offers a general view of the uses of curves and describes in some detail the principal systems that have been developed to represent and explain them, namely the Cartesian, polar and parametric systems. We shall look at how mountain paths are 'made mathematical' on a map; how planets' orbits are defined; how the most intricate objects are designed on computer; how brokers pore over market curves; and we shall study radioactive decay, the trajectory of an electron within an atom, the composition of light waves and the use of electricity. Curves are used to create studies on supply and demand, probability, population growth, the evolution of the stock market, for calculating mortgages and to determine how to launch rockets into space. The route followed by a skier, a bullet or a planet is called its trajectory. If this route is on a flat surface, it can be defined exactly by means of a mathematical formula that relates its horizontal coordinates (X) and vertical coordinates (Y) with respect to previously selected reference axes.

Reference systems: Cartesian coordinates

The first simple case we shall study is the trajectory of a skier who descends a piste in a straight line, starting from an altitude of 1,250 m.

The trajectory or downhill descent of a skier.

Naturally, a sensible skier would never follow this trajectory because they wouldn't come out of it alive, given that the final speed on this route would be 563 km/h!



Trajectory equation: y = –2 · x + 1,250 *with respect to axes X and Y selected previously.*

To determine a mathematical formula that represents this trajectory, it can be observed that:

$$\frac{300-1250}{475}=\frac{600-1250}{325}=\frac{900-1250}{175}=-2.$$

From these equations it may be deduced that each point (x,y) of the line fulfils the following relationship:

$$\frac{\text{Coordinate } y-1250}{\text{Coordinate } x}=-2.$$

Therefore, the values of the coordinates (x,y) of each point in the line shall be related mathematically by means of the formula:

$$\frac{y-1250}{x}=-2,$$

from which, isolating the coordinate y, gives:

$$y = -2x + 1{,}250.$$

This formula is called the Cartesian equation of the line.

All the straight lines on the plane X–Y have equations such as $y = m \cdot x + n$, where m is called the gradient of the line. In this case it is a constant quantity, given that it is a straight line.

It can be confirmed that this value $m = -2$ is precisely the tangent (the 'tan' button on the calculator) of the angle (116.56°) which forms the line in a positive direction on the horizontal axis (axis X).

The equation represents the descending line with respect to a system of reference coordinates – two perpendicular axes (which we call X and Y), two positive directions of the coordinates on them and a measurement unit, as indicated in the figure. This reference system is chosen before searching for the equation of any curve or straight line.

This type of coordinate system is called Cartesian in honour of René Descartes, who invented it at the beginning of the 17th century (although it was already used by a fellow contemporary mathematician, Pierre de Fermat).

To calculate m we measure the angle that forms the route of the positive direction taken on the horizontal axis (axis X) with an instrument used by topographers called a theodolite. In our case, the result is 116.56°. Therefore, the tangent of this angle is obtained by using a calculator and the result is $\tan(116.56°) = -2$.

In all other curves, the gradient is different at each point. The only 'curve' with an equal gradient at all its points is a straight line.

In the equation of a generic line, $y = m \cdot x + n$, the number n indicates the height at which the path (the line) cuts the selected vertical axis Y.

The points A, B, C, D, E highlighted are five crossing points of the skier along his straight line trajectory. Each point has its x and y coordinates indicated in brackets.

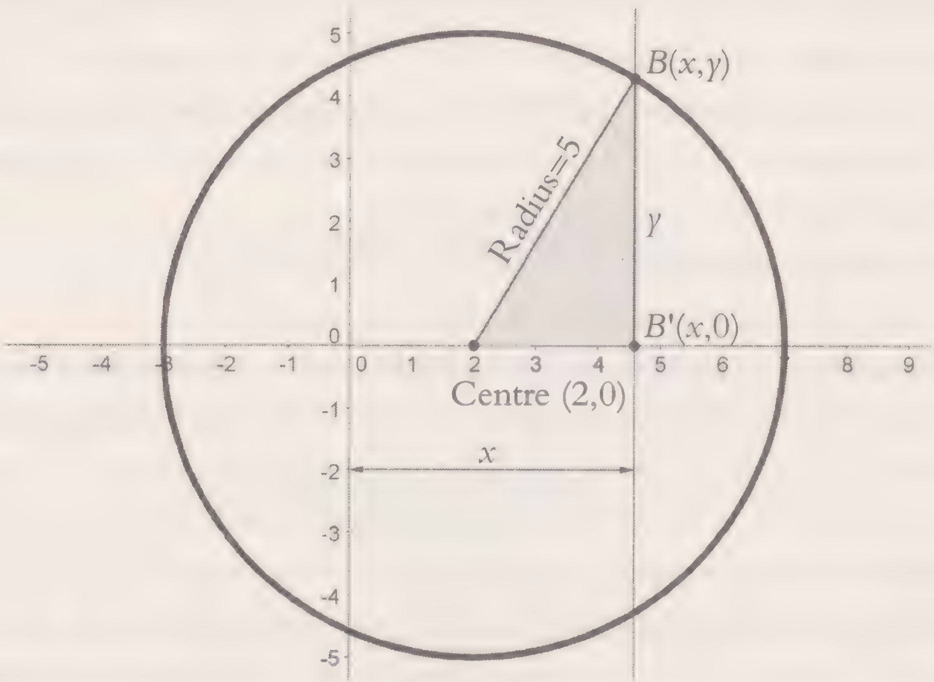
The gradient m of a line is related to the angle that forms this line with the horizontal axis or axis X, and it indicates the incline of the trajectory. This relationship is called the trigonometric tangent of the angle, and it indicates the metres the skier descends vertically for each metre that he advances horizontally. The constant number m is negative in the case of a descending trajectory, and positive in an ascending trajectory. The formula, or equation, of the skier's trajectory represents all the positions that he passes through during the descent, each one of which is indicated by its two coordinates. One could say that the equation, a simple

mathematical formula, 'contains or explains the whole curve' – its points and their characteristic form. The formula or Cartesian equation for the line $y = -2x + 1{,}250$ allows the calculation of all its positions, as can be seen in the following table:

x	Calculation of coordinate y with the formula $y = -2x + 1{,}250$	Positions of the skier (x,y)	Point
0	$-2 \cdot 0 + 1{,}250 = 1{,}250$	(0,1,250)	A
175	$-2 \cdot 175 + 1{,}250 = 900$	(175,900)	B
325	$-2 \cdot 325 + 1{,}250 = 600$	(325,600)	C
475	$-2 \cdot 475 + 1{,}250 = 300$	(475,300)	D
625	$-2 \cdot 625 + 1{,}250 = 0$	(625,0)	E

As with curves in general, the gradient changes at each of its points. For that reason it is often said that the straight line is actually the simplest curve.

The most perfect curve, well known since ancient times, is the circle. Classical Greek religion stated that the gods used it in the most important phenomena, such as the trajectories of celestial objects. It has a more complex Cartesian equation than the straight line. This equation may be deduced by applying Pythagoras' theory for a right-angled triangle:



On applying Pythagoras' theory to the triangle CBB' the result is:

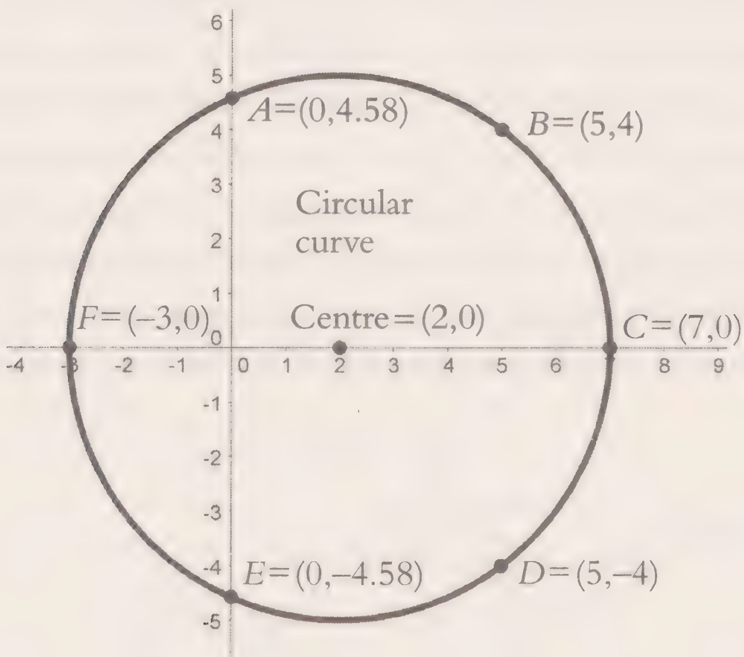
$$\text{hypotenuse } CB^2 = \text{cathetus } CB'^2 + \text{cathetus } BB'^2.$$

As the distance BC = radius = 5, the distance $BB' = y$, and the distance $CB' = x-2$, the result is: $5^2=(x-2)^2+y^2$, or: $(x-2)^2+y^2=25$.

If the variable y is resolved, the equation of the circle would be:

$$y=\pm\sqrt{25-(x-2)^2}.$$

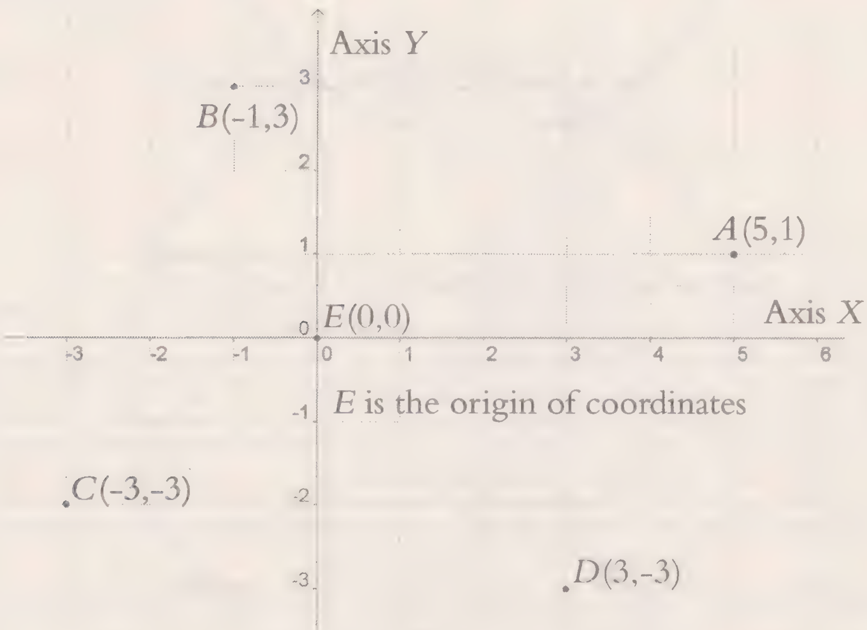
In the case of a circle with centre (2,0) and radius 5, the following graph and table of values are obtained:



x	Calculation of the coordinate and of each point, with the formula: $y=\pm\sqrt{25-(x-2)^2}$	y	Point
0	$y=+\sqrt{25-(0-2)^2}=\sqrt{21}=4.58$	4.58	A
5	$y=+\sqrt{25-(5-2)^2}=\sqrt{16}=4.00$	4	B
7	$y=+\sqrt{25-(7-2)^2}=\sqrt{0}=0.00$	0	C
5	$y=-\sqrt{25-(5-2)^2}=-\sqrt{16}=-4.00$	–4	D
0	$y=-\sqrt{25-(0-2)^2}=-\sqrt{21}=-4.58$	–4.58	E
–3	$y=-\sqrt{25-(-3-2)^2}=-\sqrt{0}=0$	0	F

The curves found on a plane may be drawn with various representation systems or systems of coordinates. The most well known is the system of Cartesian coordinates, which we have just used to write the equation of the straight trajectory of a skier and the equation of a circle. The system is also used in the game of Battleships, where two players try to destroy their opponent's fleet. They both draw the position of their ships beforehand on two sheets of paper and they determine these with a system of Cartesian coordinates. A player 'fires' at a point (giving two coordinates) in an attempt to hit one of their opponent's ships. On stating the two coordinates, the other player answers 'hit', 'sink' or 'miss'.

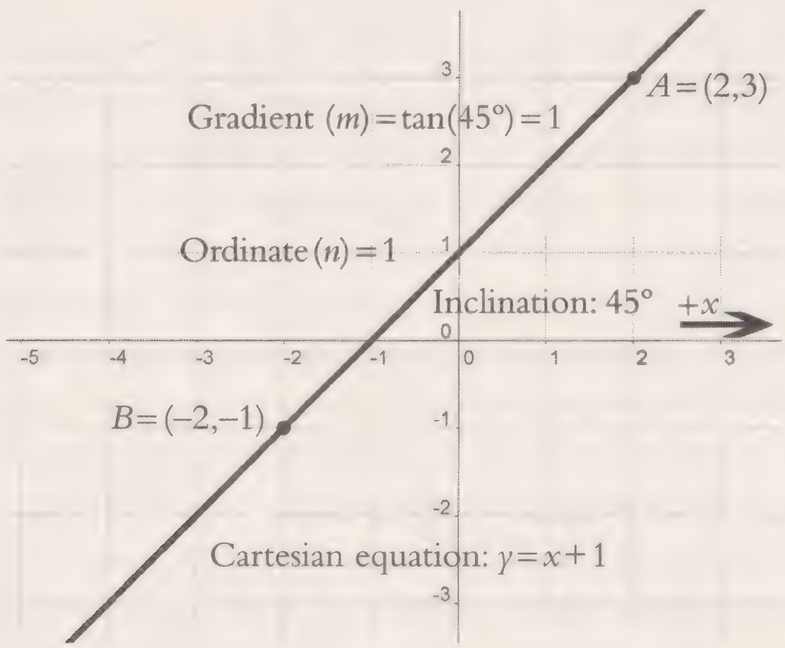
This system of plane coordinates is based on two perpendicular straight lines (called coordinate axes *X* and *Y*) upon which a series of points are marked at equal distances (measurement unit) from the intersection point of both axes, which is the initial point or coordinates' origin, (0,0). Normally, on axis *X* and from 0 towards the right the points are indicated with growing positive differentials and towards the left with decreasing negative differential points, which are called abscissa. Similarly, on the upper part of the axis *Y* points are indicated with increasing positive differentials, and on the lower part with decreasing negative differentials, called ordinates.



The points on the right-hand part of axis *X* are shown as (1,0), (2,0), (3,0),..., and the points on the left are shown as (–1,0), (–2,0), (–3,0),... Similarly, the upper and lower points of axis *Y* are (0,1), (0,2), (0,3),... and (0,–1), (0,–2), (0,–3),...Any point on the plane *X*-*Y* is indicated by its two coordinates (*x*, its abscissa, and *y*, its

ordinate). For example, the point (5,1) indicates that this point is located at a distance of 5 units from axis *Y*, and a distance of 1 unit from axis *X*, as shown in the previous figure.

The two axes divide the plane into four areas called quadrants. Point *A* is located in the first quadrant; *B* is in the second; *C* is in the third; and *D* in the fourth. Therefore, the references necessary to define a system of Cartesian coordinates are *two* perpendicular axes (the abscissas and ordinates, X and Y), with the positive and negative directions indicated and *one* measurement unit to determine the distance between two consecutive coordinates. As the following illustration demonstrates, if a straight line is drawn through the points *A* (2,3) and *B* (–2,–1) on a plane that already indicates the coordinate axes *X* and *Y*, we can see that all the points on this line have the same relationship between its *Y* and *X* coordinates.



In this case, this mathematical relationship is expressed as follows: the value of *Y* of each point always corresponds to the value of *X* plus 1.

It may be considered that the mathematical expression (equation or formula) $y=x+1$ represents all the points of the line drawn and contains, to a certain extent, the entire line and its form. This formula is called the equation of the line.

The values that x and y can take are called variables (because they may have a lot of values). In the equations expressed in the Cartesian system, *x* is known as an independent variable, and y, as a dependent variable or variable function.

This means that variable x may adopt a set of values at will (any that the person drawing the curve wants), but the values of y 'depend' on the result that is obtained for each value of x selected in the equation (or formula) of the specific curve.

Comparing the equation of this line with the more general equation of a line in Cartesian coordinates, $y = m \cdot x + n$ seen previously, in this case the values of m and n are: $m = 1$, $n = 1$. The angle that forms the line with axis X is the angle or incline of the tangent of which is 1. And this corresponds to an angle of 45°, as can be seen in the graph. The value $n = 1$ indicates that the crossing point of the line with axis Y is the point (0,1).

All the points of this line have its coordinates in the relationship $(x, x+1)$, as in (–3,–2), (–2,–1), (–1,0), (0,1), (1,2), (2,3),... In this case, the selected values of the independent variable x are: –3, –2, –1, 0, 1, 2, and the 'dependent' values of y that correspond to them are: –2, –1, 0, 1, 2, 3.

Evidently, the person drawing the line can give many more values to x, calculating those that correspond to y. However, in the case of straight lines only two points are necessary, given that a straight line joins only two points.

If at this juncture one considers the circle drawn previously with a radius of 5 and centre of (0,2), it may be confirmed that its equation in Cartesian coordinates is: $y = \pm\sqrt{25-(x-2)^2}$. The drawing of this circle requires knowledge of more than two points to establish its form and position. The sign + of the root corresponds to the points of the upper part, and the sign – to the points of the curve in the lower part of axis X. All the points of the upper part of this circle have the coordinates $(x, +\sqrt{25-(x-2)^2})$, and those of its lower part, $(x, -\sqrt{25-(x-2)^2})$.

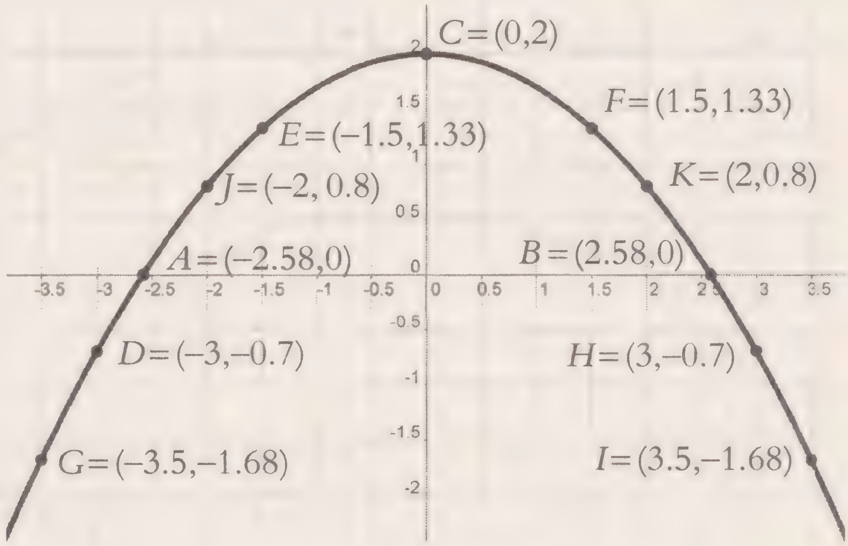
As has been seen, all the straight lines have first-degree equations, which means that in the equation the x and the y are raised to the first power (x^2, y^2 or higher powers do not appear). It is said that all such lines form the family of first-degree curves.

In contrast, the equation of the circle written with the y kept separate is $y = \pm\sqrt{25-(x-2)^2}$. (This type of Cartesian equation is called 'explicit' because the variable y is isolated to the left of the equation.) In an implicit form (in which neither of the two variables are isolated, x and y) it is $(x-2)^2 + y^2 = 25$. In equations for these types of curves, powers of the variables x or y appear that are superior to one. For this reason they say that the circle and other similar curves such as the parabola, hyperbola and ellipse, belong to the family of second-degree curves. (They are also know as conical, and we shall see why further on.)

The parabolae have Cartesian equations of the type $y = ax^2 + bx + c$; for example, $y = -0{,}3x^2 + 2$, or $y = 3x^2 - 2$.

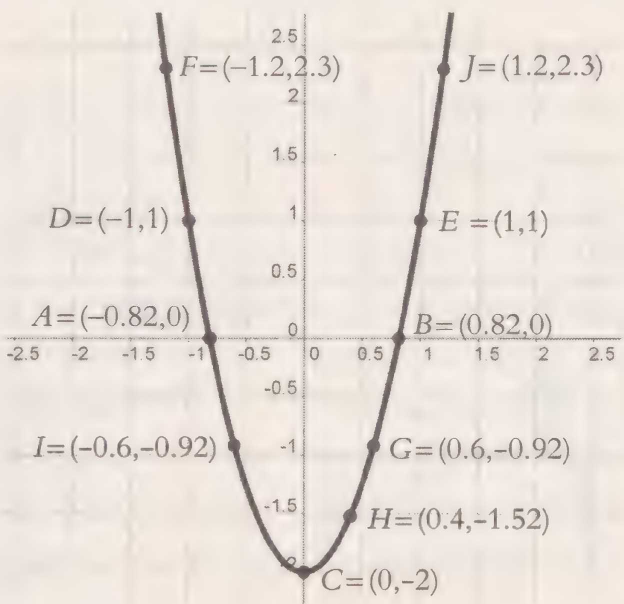
As is demonstrated in the following graphs, the two parabolae have a symmetry axis in the form of axis *Y*. In the first case, the arms curve downwards, and in the second case, they curve upwards. The two arms are more 'closed' in the parabola $y = 3x^2 - 2$ and more "open" in the parabola $y = -0.3x^2 + 2$. These two curves have been drawn calculating different points of these curves starting from its equation.

x	*y*	Point
–3.5	–1.68	*G*
–3	–0.7	*D*
–2.5	0.13	close to *A*
–2	0.8	*J*
–1.5	1.33	*E*
0	2	*C*
1.5	1.33	*F*
2	0.8	*K*
2.5	0.13	close to *B*
3	–0.7	*H*
3.5	–1.68	*I*



Parabola curve with vertical symmetry with Cartesian equation $y = -0.3x^2 + 2$.

x	*y*	Point
–1.20	2.32	*F*
–1.00	1	*D*
–0.80	–0.08	close to *A*
–0.60	–0.92	*I*
0.00	–2	*C*
–0.40	–1.52	*H*
0.60	–0.92	*G*
0.80	–0.08	close to *B*
1.00	1	*E*
1.20	2.32	*J*



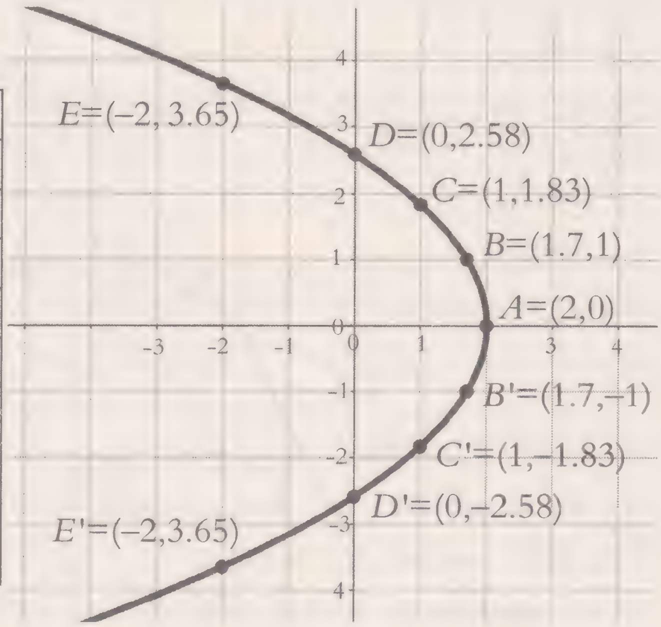
Parabola curve with vertical symmetry with Cartesian equation: $y = +3x^2 - 2$.

Other types of parabolic curves exist with equations similar to those above but with interchanged variables of x and y. Two examples would be $x=-0.3y^2+2$; $x=3y^2-2$. If the variable y is removed it would remain as follows:

$$y=\pm\sqrt{\frac{-x+2}{0.3}}\ ;\quad y=\pm\sqrt{\frac{x+2}{3}}.$$

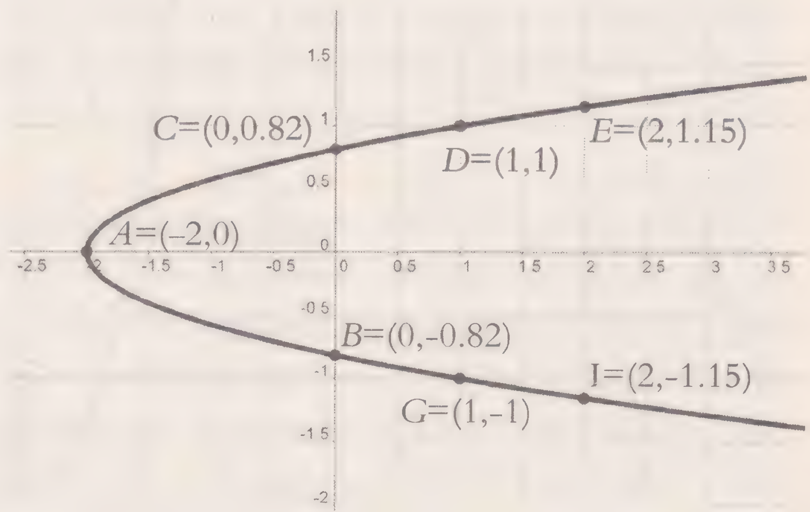
The symmetry axis of these two parabolae is X, and their graphs are the following:

x	y	Point
−2	3.65	*E*
0	2.58	*D*
1	1.83	*C*
1.7	1	*B*
2	0	*A*
1.7	−1	*B'*
1	−1.83	*C'*
0	−2.58	*D'*
−2	−3.65	*E'*



Parabola curve with horizontal symmetry. Cartesian equation $x=-0.3y^2+2$, *or rather* $y=\pm\sqrt{\frac{-x+2}{0.3}}$.

x	y	Point
2	1.15	*E*
1	1	*D*
0	0.82	*F*
−2	0	*E*
0	−0.82	*B*
1	−1	*G*
2	−1.15	*I*



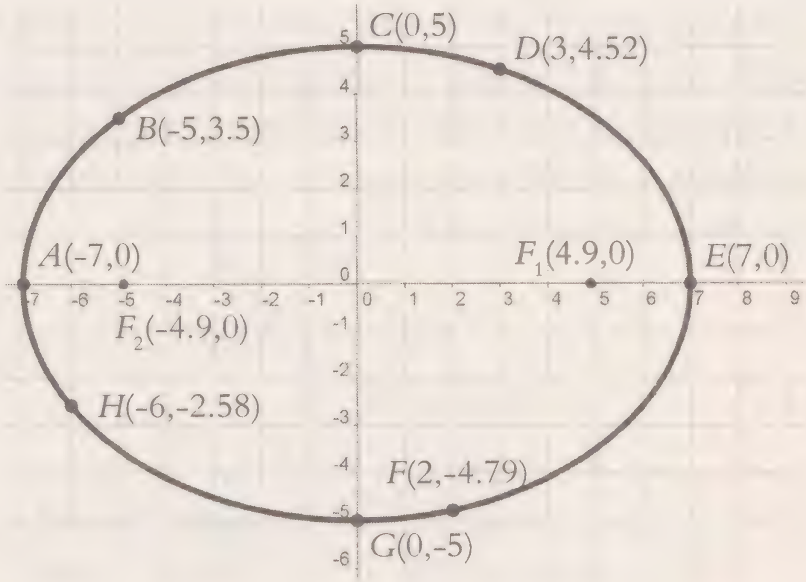
Parabola curve with horizontal symmetry. Cartesian equation $x=3y^2-2$, *or rather* $y=\pm\sqrt{\frac{x+2}{3}}$.

The ellipse is another curve of the family of second-degree curves. Its Cartesian equation can be expressed as

$$\frac{x^2}{a^2}+\frac{y^2}{b^2}=1,$$

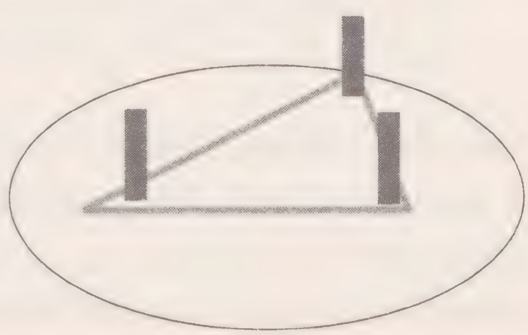
in which a and b are the lengths of half its greatest diameter and its smallest diameter. In the following graph we can see an ellipse with a greater diameter of 14 and smaller diameter of 10. Their focal points are F_1 and F_2. The basic property of the ellipse is that the sum of the distances of any of its points to the two focal points is constant, and its value is $2a$. In the case of the ellipse that is indicated as follows, this sum is: $2a=2\cdot 7=14$.

x	y	Point
–7	0	A
–5	3.5	B
0	5	C
3	4.52	D
7	0	E
2	–4.79	F
0	–5	G
–6	–2.58	H



Ellipse curve. Cartesian equation: $\frac{x^2}{49}+\frac{y^2}{25}=1$, *or rather* $y=\pm 5\sqrt{(1-\frac{x^2}{49})}$.

The ellipse is curve that has been well known since the time of Classical Greece. It is known as the 'gardener's curve' because gardeners draw ellipses in the soil for planting flowers. They achieve this curve by using a cord and two poles driven into the soil at the focal points of the ellipse, as demonstrated in the illustration. The fixed points are the focal points of the ellipse.

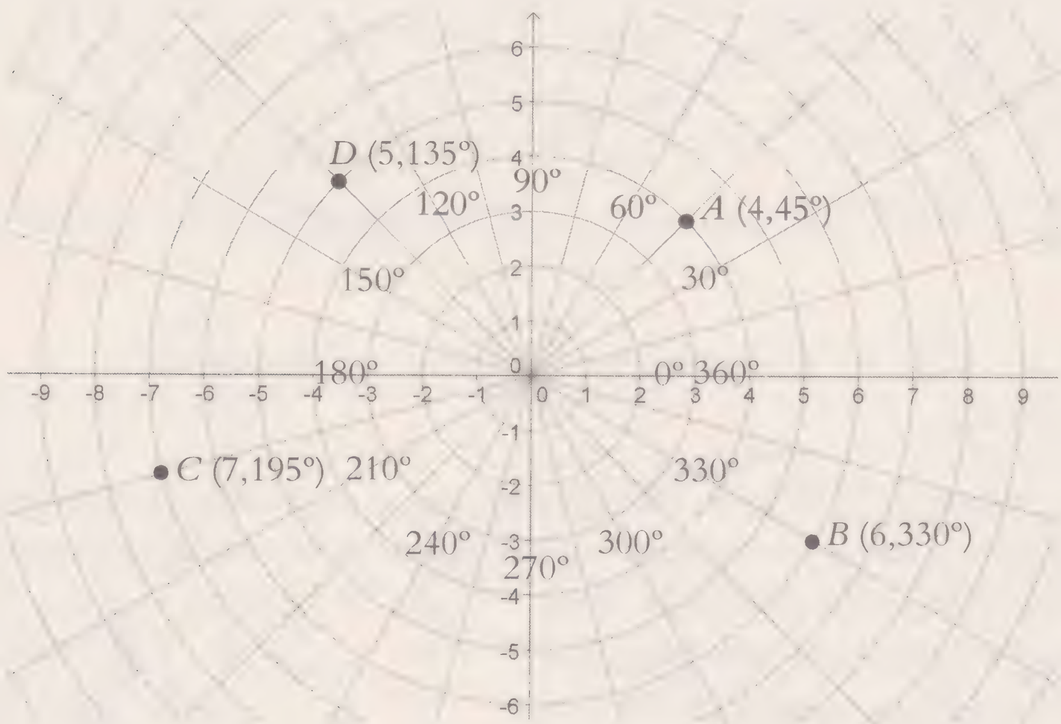


Civilisation has gone a long way since the curves that followed planets and stars

were thought to be circles. These were considered to be the most perfect curve and, therefore, the gods were said to have chosen them for the trajectories of all heavenly objects. Naturally, the centre of all those circles had to be the Earth. The first person who dared to challenge this hypothesis was Johannes Kepler (1571-1630), who arrived at the conclusion that the planetary curves were elliptic.

Alternative systems to Cartesian coordinates have been defined to represent curves, such as polar coordinates and parametric coordinates. The concept of the system of polar coordinates was defined by Isaac Newton (1642-1727) in his first book *Method of Fluxions* (written in 1671 and published in 1736), where he invented the concept of derivatives – although this method had already been used by the Swiss scientist Jakob Bernoulli (1654-1705) in several works for the scientific journal *Acta eruditorum* in 1690.

To define a system of polar coordinates, the references taken are a point of origin (pole) and an axis from which angles are measured. These are considered to be positive if they rotate anticlockwise. Four points and their polar coordinates are shown in the following diagram. One is positioned in each quadrant: $A\,(4,45°)$, $B\,(6,330°)$, $C\,(7,195°)$ and $D\,(5,135°)$.



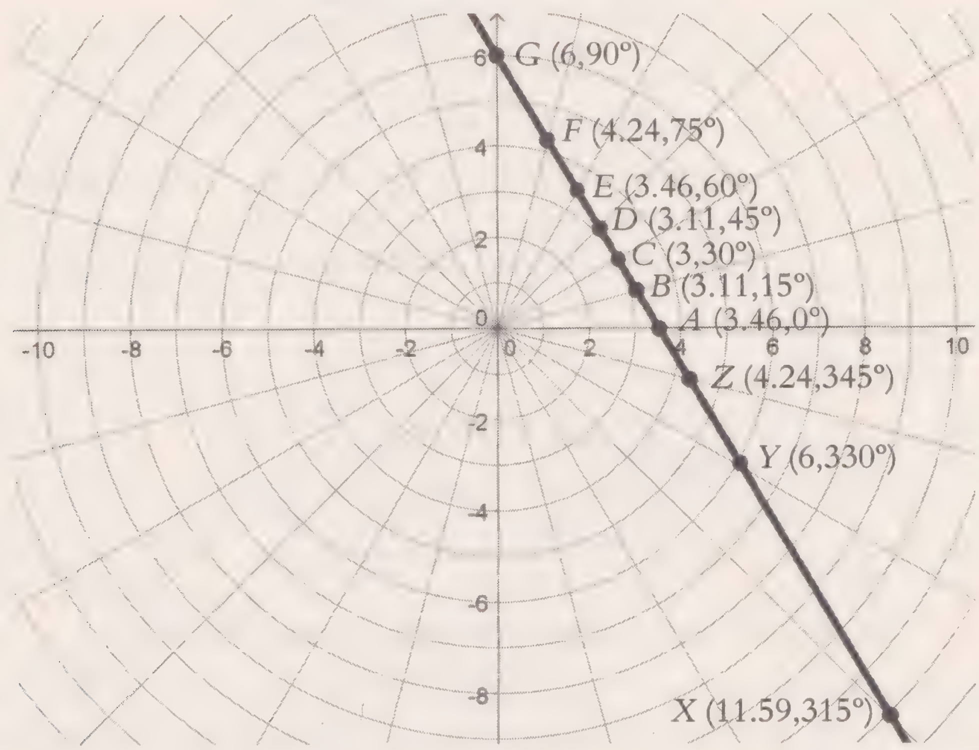
Instead of both being lengths as in Cartesian coordinates, a point in a polar system is described as one length and one angle. The length is the distance from the point to the pole. The angle is formed by the segment that connects to the pole and the

polar axis taken in the positive direction of rotation. In general, the letter r and the Greek letter θ are used; in other words, a polar point is indicated as $P(r,\theta)$.

The equation or formula of a curve can also be expressed in polar coordinates. In the case of a line such as the one in the diagram, which has a Cartesian equation of $y=-1.73x+6$, the equation in polar coordinates is

$$r=\frac{3}{\cos(\theta-30)}.$$

Point	A	B	C	D	E	F	G	...	...		...	X	Y	Z	A
Angle	0	15	30	45	60	75	90	105	120	...	300	315	330	345	360
Radius	3.46	3.11	3.00	3.11	3.46	4.24	6.00	11.59	∞		∞	11.59	6.00	4.24	3.46

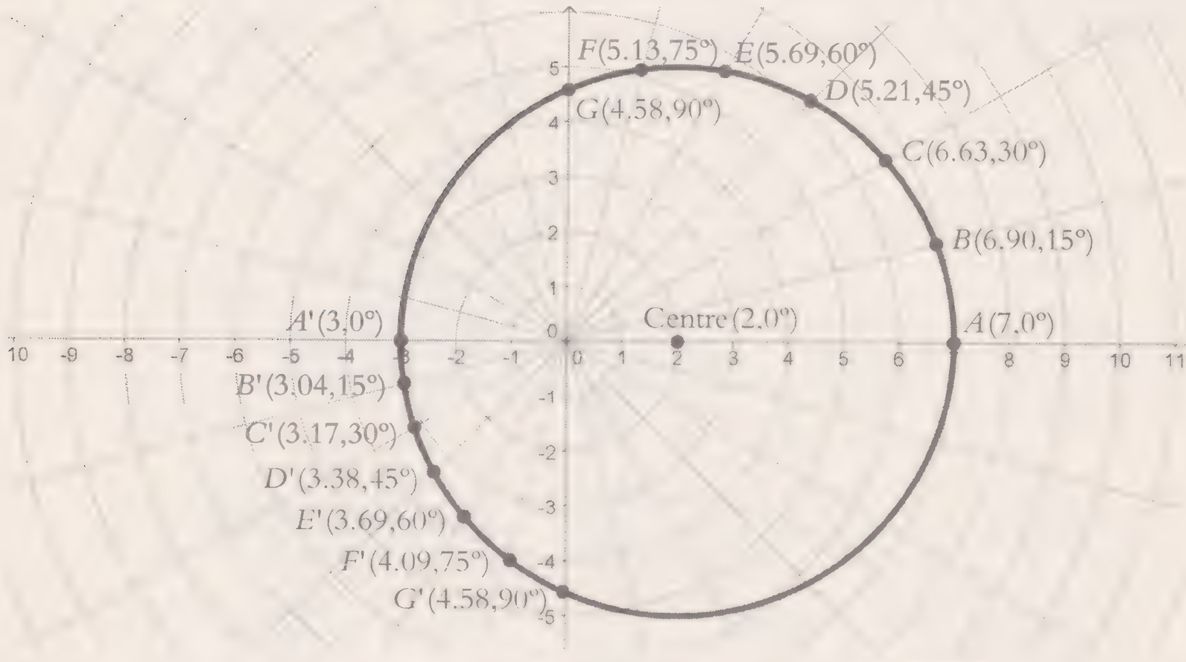


In the case of the circle of radius 5 and centre (2,0) described previously with a Cartesian equation $y=\pm\sqrt{25-(x-2)^2}$, its equation in polar coordinates is:

$$r=2\cos\theta\pm\sqrt{4\cos^2\theta+21}.$$

The position and the polar coordinates of 14 of its points are highlighted in the accompanying graph and table: r_1 corresponds to the plus sign of the equation, and r_2, to the minus sign for the same angle θ.

Angle θ	r_1	Point	r_2	Point
0	7.00	A	–3.00	A'
15	6.90	B	–3.04	B'
30	6.63	C	–3.17	C'
45	6.21	D	–3.38	D'
60	5.69	E	–3.69	E'
75	5.13	F	–4.09	F'
90	4.58	G	–4.58	G'



Equation of the circle in polar coordinates: $r = 2\cos\theta \pm \sqrt{4\cos^2\theta + 21}$.

The ellipse, with its graph produced previously from its equation in Cartesian coordinates,

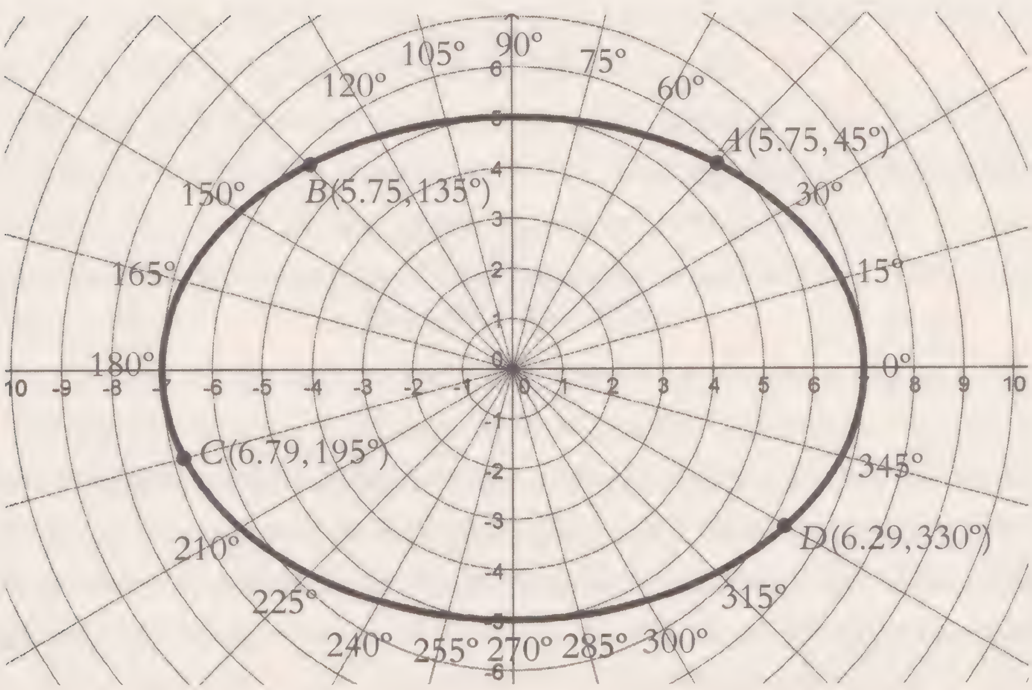
$$\frac{x^2}{49} + \frac{y^2}{25} = 1,$$

may also be expressed in polar coordinates:

$$r = \frac{35}{\sqrt{49\sin^2\theta + 25\cos^2\theta}},$$

and drawn with a polar representation system:

Angle θ	Point	r
0	...	7.00
15	...	6.79
30	...	6.29
45	*A*	5.75
	....	
135	*B*	5.75
150	...	6.29
165	...	6.79
180	...	7.00
195	*C*	6.79
	....	
330	*D*	6.29
345	...	6.79
360	...	7.00



On comparing the equations of the same curve in Cartesian and polar systems, it can be seen that their aspects are very different, and they also present different complexities.

The equation of the line in Cartesian coordinates,

$$y=-1.73x+6,$$

results in polar coordinates:

$$r=\frac{3}{\cos(\theta-30)}.$$

The equation of the circle in Cartesian coordinates,

$$y=\pm\sqrt{25-(x-2)^2},$$

results in polar coordinates:

$$r=2\cos\theta\pm\sqrt{4\cos^2\theta+21}.$$

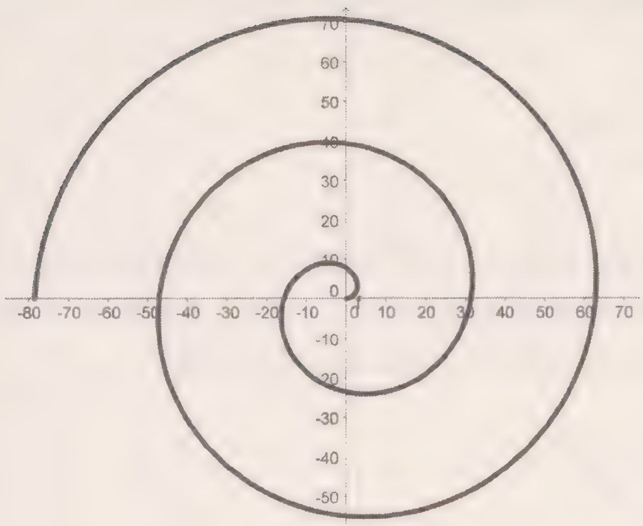
The equation in Cartesian coordinates of the ellipse centred on the origin,

$$\frac{x^2}{49}+\frac{y^2}{25}=1,$$

results in polar coordinates:

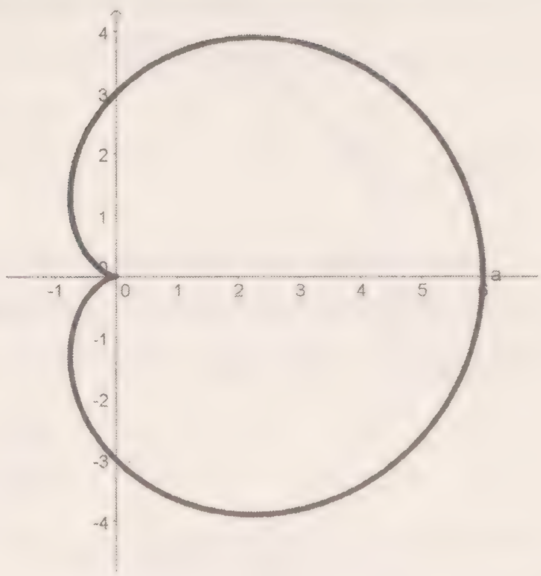
$$r=\frac{35}{\sqrt{49\sin^2\theta+25\cos^2\theta}}.$$

In these three cases, the two types of equations (Cartesian and polar) of lines, circles and ellipses appear to be equally complicated. However, in other curves, the difference in the complexity of its equations is enormous depending on the coordinates system chosen. This is the case for lemniscate and Archimedes spiral equations, as demonstrated in the following figures:



Archimedes spiral. Equation in polar coordinates: $r=5\theta$. *Implicit equation in Cartesian coordinates:*

$$y - x\tan\left(\frac{\sqrt{x^2+y^2}}{5}\right) = 0.$$



Cardioid. Equation in polar coordinates: $r=3(1+\cos\theta)$.
Implicit equation in Cartesian coordinates: $(x^2+y^2)^2-6x(x^2+y^2)-9y^2=0$.

Another well-used type of curve representation system is called parametric. This system is more sophisticated and stem from Euler and Gauss (18th and 19th centuries), who used it widely in their studies of curves and surfaces. Parametric coordinates have the advantage of being intrinsic, meaning that they don't depend on external axes, as do both Cartesian or polar coordinates. Instead they are based on a reference system included in the object or curve itself.

Parametric coordinates are used to determine a position on the Earth (longitude and latitude). They are two parameters that do not refer to a Cartesian coordinate system external to our planet, which would be three-dimensional, but rather to imaginary circles positioned on the Earth's surface itself, which can be considered two-dimensional.

The number of parameters required to describe a geometric object (a flat curve or a 3D entity) indicates its dimension. Flat lines and curves have one dimension. Therefore, their expression requires only one parameter.

However, to be able to represent a flat curve graphically it is necessary to transform the parametric coordinates into Cartesian coordinates. This leads to two equations that allow us to calculate the x and y of each point based on the parameter, generally indicated by the letters u or t. These equations are parametric.

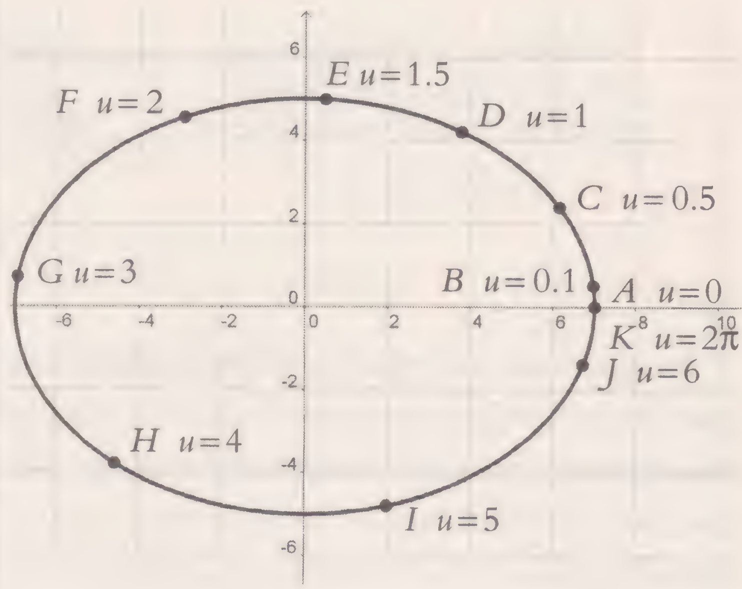
The parametric equations of the ellipse (of semi-axes 7 and 5) with its centre on the origin (of the coordinates) are:

$$\begin{cases} x = 7\cos(u) \\ y = 5\ \sin(u) \end{cases}$$

The Cartesian coordinates of all its points can be obtained by assigning values to the parameter u between 0 and 2π radians ($\cong 6.28$), which are the equivalent of one complete turn (from 0° to 360°). A radian is a measurement of angle

$$\frac{360}{2\pi} \cong 57.30°.$$

u	$x(u)$	$y(u)$	Point
0	7.00	0.00	A
0.1	6.97	0.50	B
0.5	6.14	2.40	C
1	3.78	4.21	D
1.5	0.50	4.99	E
2	–2.91	4.55	F
3	–6.93	0.71	G
4	–4.58	–3.78	H
5	1.99	–4.79	I
6	6.72	–1.40	J
2π	7.00	0.00	K



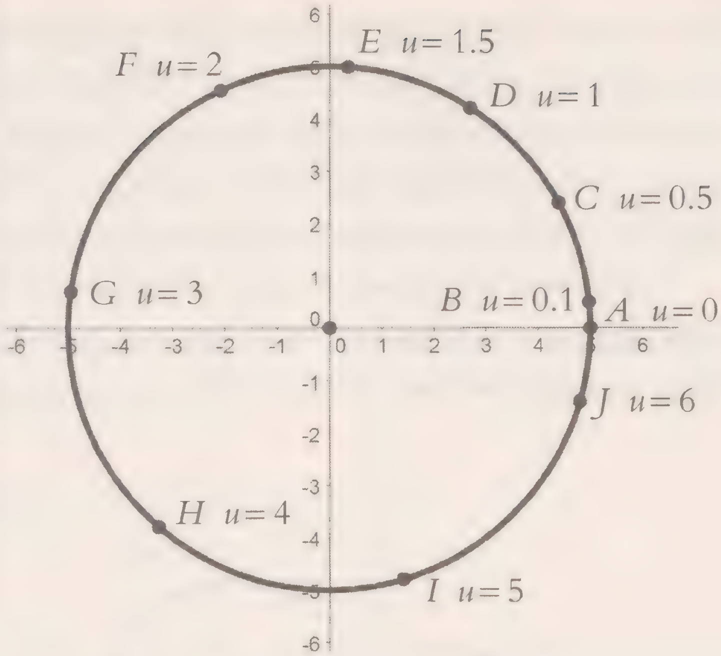
Ellipse curve. Parametric equations: $\left.\begin{array}{l} x=7\cos(u) \\ y=5\sin(u) \end{array}\right\}$

The parametric equations of the circle of radius 5 and centre on the origin (0,0) are:

$$\begin{cases} x = 5\cos(u) \\ y = 5\ \sin(u) \end{cases}$$

The Cartesian coordinates of all its points can be obtained from the values assigned to the parameter u between 0 and 2π radians ($\cong 6.28$), which are the equivalent of one complete turn.

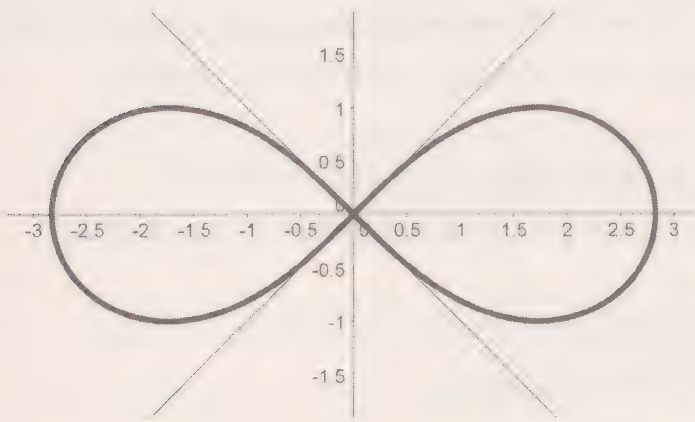
u	x(u)	y(u)	Point
0.00	5.00	0.00	A
0.10	4.98	0.50	B
0.50	4.39	2.40	C
1.00	2.70	4.21	D
1.50	0.35	4.99	E
2.00	–2.08	4.55	F
3.00	–4.95	0.71	G
4.00	–3.27	–3.78	H
5.00	1.42	–4.79	I
6.00	4.80	–1.40	J
2π	5.00	0.00	A



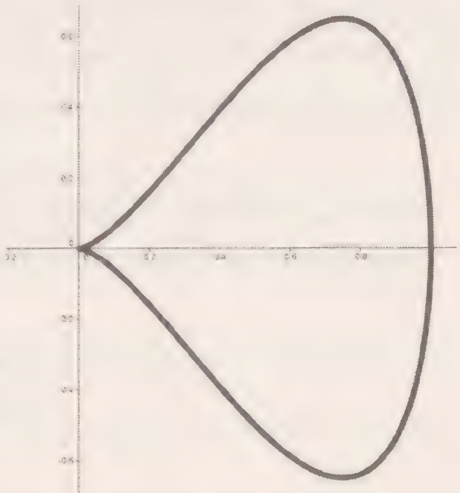
Parametric equations: $\left.\begin{aligned} x &= 5\cos(u) \\ y &= 5\sin(u) \end{aligned}\right\}$

Curves for computer-drawn objects

The forms of very diverse natural or artificial objects can be constructed with mathematical curves, such as the infinity symbol or a pear shape.



Cassini oval. Equation $y=\pm\sqrt{\sqrt{16x^2+16}-x^2-4}$ *in Cartesian coordinates.*

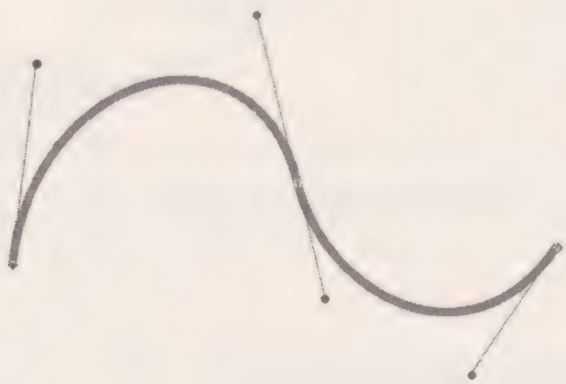


Piriform quartic curve. Equation $y=\pm 2x\sqrt{x-x^2}$ *in Cartesian coordinates.*

The parametric equations of the curves explained previously are used to a great extent in the generation of graphics on a computer (using Computer Assisted Design, or CAD, programs). To draw straight sections with a CAD program, you have to click

on the drawing tool (generally an icon shaped like a pen), then move the mouse and click on a new point. To draw lines that are slightly curved, you click and keep the button pressed down while you adjust the shape of the curve. When you let go of the mouse you have established the point or points of control. The initial shape that has been drawn can be modified afterwards by moving these control points.

The parametric equations for CAD curves, called Bézier curves, date back to 1962 and were defined by the French engineer Pierre Bézier, who used them to design automobile parts for Renault.



Bézier curve with three control points.

There are different types of Bézier curves. Linear Bézier curves are defined by two points, A and B. A linear Bézier curve is a straight line that passes through these points and is expressed by the mathematical parametric formula or equation:

Bezier_CURVE_1 $(u)=(1-u)\,A+uB$ for values of u between 0 and 1,
which can be developed in two parametric equations such as the following:

$$\begin{cases} x(u)=(1-u)\cdot x_A+u\cdot x_B \\ y(u)=(1-u)\cdot y_A+u\cdot y_B \end{cases}$$

If the crossing points are A (0,0) and B (3,2), then the equation of the line Bezier_CURVE_1 is one of a line such as:

$$\begin{cases} x=(1-u)\cdot 0+u\cdot 3 \\ y=(1-u)\cdot 0+u\cdot 2 \end{cases} \text{ or: } \begin{cases} x=3u \\ y=2u \end{cases} \text{ or: } y=\frac{2}{3}x,$$

which is a line that passes through the origin (0,0) and has a 2/3 gradient.

To draw a curve with a CAD program that is determined by three points (two crossing and one control) quadratic Bézier curves (or second-degree parabolae) are used. Given three points A (x_A,y_A), B (x_B,y_B) and C (x_C,y_C), a second-degree curve can be determined which passes through the points A and C with the control condition that they have straight tangents that are precisely the lines AB and BC (B is the so-called control point).

The curve that Bézier proposed which complies with these conditions is expressed in parametric coordinates with two equations:

$$\begin{cases} x(u) = (1-u)^2 \cdot x_A + 2u(1-u) \cdot x_B + u^2 \cdot x_C \\ y(u) = (1-u)^2 \cdot y_A + 2u(1-u) \cdot y_B + u^2 \cdot y_C \end{cases}$$

The drawing process starts at point A, moving towards B (control point), in the direction AB, although the curve does not pass through B, and ends at C, but the drawing of the curve heads in the direction of BC.

Mathematically this means that AB is a line that is a tangent to the curve at point A, and BC is a tangent to the curve at point C. For values of the parameter u between 0 and 1, the desired inflection points (where there is a change in direction) can be calculated to draw a curve that complies with the stated conditions.

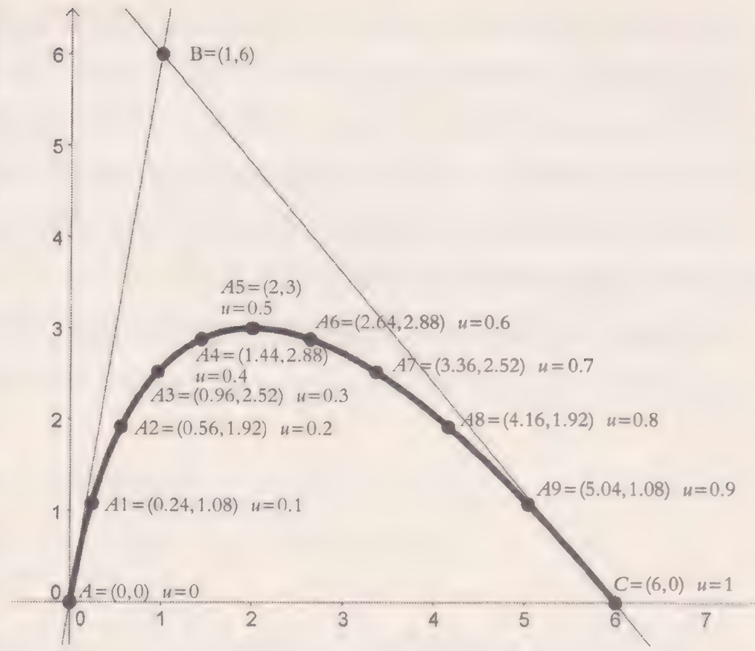
For inflection points A (0,0) and C (6,0) and control point B (1,6), the equation of the line Bezier_CURVE_2 is developed in two parametric equations that arise in a second-degree parabolic curve:

$$\begin{cases} x(u) = (1-u)^2 \cdot 0 + 2u(1-u) \cdot 1 + u^2 \cdot 6 \\ y(u) = (1-u)^2 \cdot 0 + 2u(1-u) \cdot 6 + u^2 \cdot 0 \end{cases}$$

$$\text{or: } \begin{cases} x(u) = 4u^2 + 2u \\ y(u) = -12u^2 + 12u \end{cases}$$

Therefore the computer assigns values of u between 0 and 1, according to the necessary points that make up the intended drawing, calculating the coordinates of the points of the Bézier curve and 'marking' them on the screen. This allows the number of points to increase when you zoom into the design. In the example illustrated on the following page, the computer has calculated 11 inflection points on the curve, increasing the parameter u in values of tenths.

u	*x(u)*	*y(u)*
0	0	0
0.1	0.24	1.08
0.2	0.56	1.92
0.3	0.96	2.52
0.4	1.44	2.88
0.5	2	3
0.6	2.64	2.88
0.7	3.36	2.52
0.8	4,16	1.92
0.9	5.04	1.08
1	6	0



Bézier curve (parabola) for inflection points A*(0,0),* C*(6,0) and control point* B*(1,6).*

Parametric equations $\left.\begin{aligned} x(u) &= 4u^2 + 2u \\ y(u) &= -12u^2 + 12u \end{aligned}\right\}$

This Bézier curve expressed in parametric coordinates gives rise – after a series of operations with symbolic calculus software, such as Derive – to an equation in Cartesian coordinates which is a second-degree parabola with the accompanying Cartesian equation between $x=0$ and $x=6$:

$$y=\frac{3}{2}(3\sqrt{4x+1}-2x-3),$$

in the explicit form (with the y separate), or in the implicit form: $9x^2+6xy-54x+y^2+9y=0$.

Curves that define physical and chemical phenomena

Many physical phenomena can be represented with mathematical curves through their formulae or equations expressed with the coordinate systems described thus far. Examples include some of the trajectories that an electron may take within an atom, which have the form $\cos(n\theta)$, with n always even. For an odd n, we have, for example, a curve called a polar rose, studied by Guido Grandi, whose equation in polar coordinates is $r=\cos(5\theta)$.

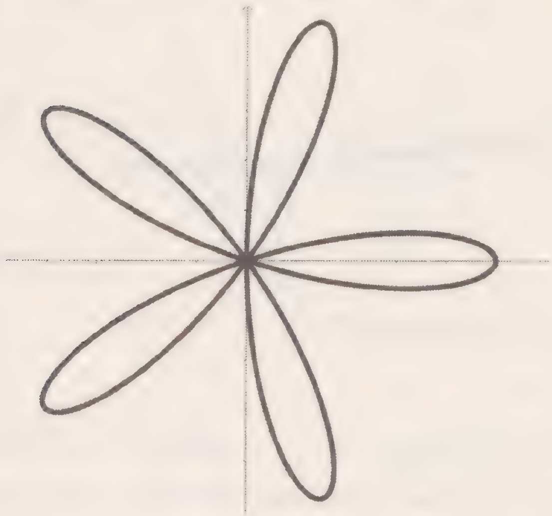
Another very attractive curve is the trajectory of a particle upon which two harmonious movements act simultaneously (such as that of a pendulum), moving

at right angles. Other examples include the movement of an object elongated at the same time by two different springs that act in perpendicular directions, or the wave resulting from the composition of two waves of different frequencies that collide in two perpendicular directions.

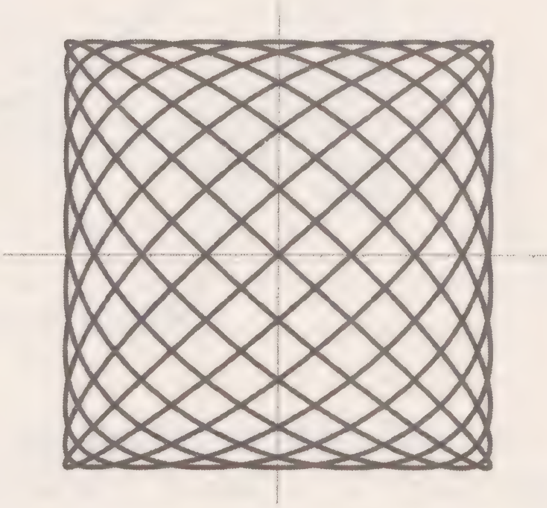
The final trajectory of the particle or the resulting wave produces so-called Lissajous curves, an example of which can be seen in the following illustration and which corresponds to the parametric equations:

$$\begin{cases} x = \sin(10u) \\ y = \sin(9u) \end{cases}$$

In this case the variable u is time.



Trajectory of an electron, Guido Grande rose curve. Equation in polar form: $r = \cos(5\theta)$.



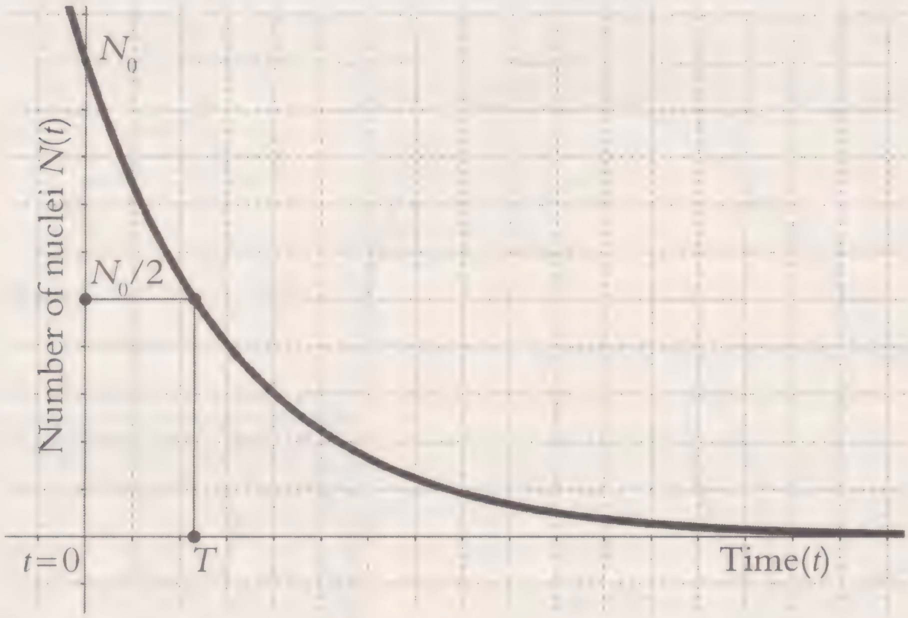
Lissajous curves. Parametric equations:

$$\begin{cases} x = \sin(10u) \\ y = \sin(9u) \end{cases}$$

The nuclei of atoms are composed of protons and neutrons, which are held together by the strong nuclear force. Some nuclei have a combination of protons and neutrons that does not lead to a stable configuration. Consequently they are unstable or radioactive and to arrive at a stable configuration they emit certain particles. The different types of radioactive decay are classified according to the type of particles emitted: alpha, beta or gamma.

Experimentally it has been observed that all the simple radioactive processes follow a curve that decreases exponentially. If N_0 is the number of radioactive nuclei that there were at the initial time, after a certain time t, the number of radioactive nuclei present, N, has reduced.

The number of radioactive nuclei present at the time t counted from $t=0$ is $N=N_0e^{-kt}$, where k is a characteristic of the radioactive substance called the constant of disintegration.



Decreasing exponential curve $N=N_0e^{-\lambda}$.

For each radioactive substance there is a fixed interval $T_{1/2}$, called the half-life, after which the number of its initial nuclei is reduced by half. Putting it into an equation

$$N=\frac{N_0}{2},$$

we obtain

$$\frac{N_0}{2}=N_0e^{-\lambda T},$$

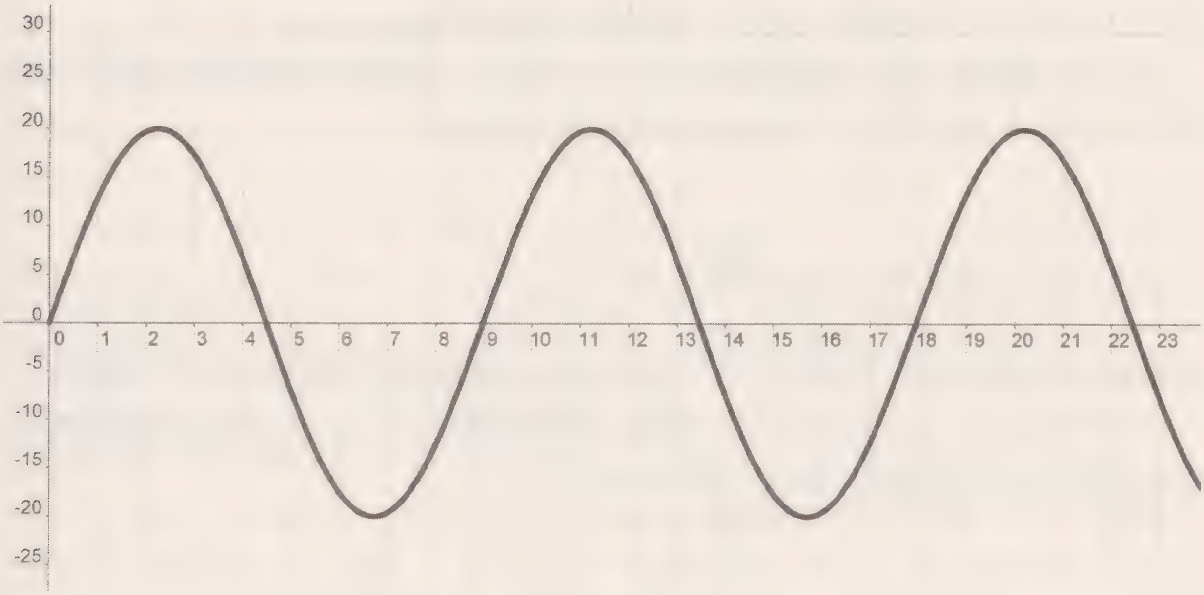
which gives:

$$T_{½}=\frac{\ln 2}{\lambda},$$

which allows us to calculate the half-life $T_{1/2}$ from the disintegration constant λ. For uranium 238, $T_{1/2}=4{,}468\cdot 10^9$ years.

Alternating electric current is explained as a periodic curve that indicates the intensity of the current according to time, in other words the number of charges

per second that are moved through the conductor. Its magnitude and direction vary periodically between certain maximum and minimum values. It is the common mode of distributing energy generated from solar, wind, hydraulic, nuclear and thermal sources, among others. The alternating wave of current most often takes the form of a sinusoidal curve, as this produces a more efficient transmission of energy. Its equation is $I = I_{max} \sin(t)$. However, in certain electronic applications other types of periodic curves are used, such as the cubic or quadratic curves.



Sinusoidal curve y=sin(2π/9x).

Alternating current tends to be the form in which electrical power arrives at homes and offices, although audio and radio signals transmitted through electric cables are also examples of an alternating current. In these cases, the end point tends to be the transmission and recuperation of coded information over the alternating current signal.

The frequency of alternations between the maximum and minimum in the alternating current curve is from 50 Hz (50 times per second) in Europe and 60 Hz in North America.

Market curves

Curves are also used in social phenomena. In some cases these curves have a mathematical equation which permits us to draw them precisely in advance. In other cases, they are curves constructed with empirical data (taken from reality).

Curves have a very important application in market analysis, for example to check the usage of a detergent that has been estimated to be used in 26% of homes. To confirm the reliability of this percentage, a questionnaire is presented to 12 households asking about its use.

Supposing that it is true that it is used in 26% of homes in the country, we want to calculate the probability of finding a between 6 and 9 users of this brand from the 12 household sample.

In this case, it is a variable with only two values ('yes' they use it, 'no' they don't). The estimated probability of yes is 26%, and no is 74%.

Using the laws of combinatorics we can deduce that the probability $PB(k)$ that there are k users of the brand in a sample of n users, is:

$$PB(k) = \binom{n}{k} p^k (1-p)^{n-k},$$

where $p=0.26$ and $(1-p)=1-0.26=0.74$ (the probabilities are many to one).

This formula corresponds to a binomial probability function designed by Jakob Bernoulli in the 17th century. The formula

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

denotes the number of unique combinations that can be obtained from n objects in groups of k objects.

From the laws of probability it is deduced that the probability that one or more of the possible events occurs is calculated by adding the individual probabilities for each of these, provided they are independent of each other.

In this case, the probability that six users appear in our sample is:

$$PB(6) = \binom{12}{6} 0.26^6 \cdot 0.74^6 = 0.0468708012.$$

Applying the formula above produces the following table, calculated in Excel, of the values $PB(6)$ to $PB(9)$.

k	6	7	8	9
$PB(k)$	0.0468708012	0.0141155309	0.0030996943	0.0004840363

The probability distribution graph is presented as a low variable function with the formula:

$$PB(k) = \binom{n}{k} p^k (1-p)^{n-k},$$

with $n = 12$ and $k = 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12$. The requested probability, which corresponds to the values between 6 and 9 is:

$$\begin{gathered} \mathrm{PB}(6 \leq x \leq 9) = \mathrm{PB}(6) + \mathrm{PB}(7) + \mathrm{PB}(8) + \mathrm{PB}(9) = \\ = 0.0468708012 + 0.0141155309 + 0.0030996943 + 0.0004840363 = \\ = 0.0645700627 = 6.46\%. \end{gathered}$$

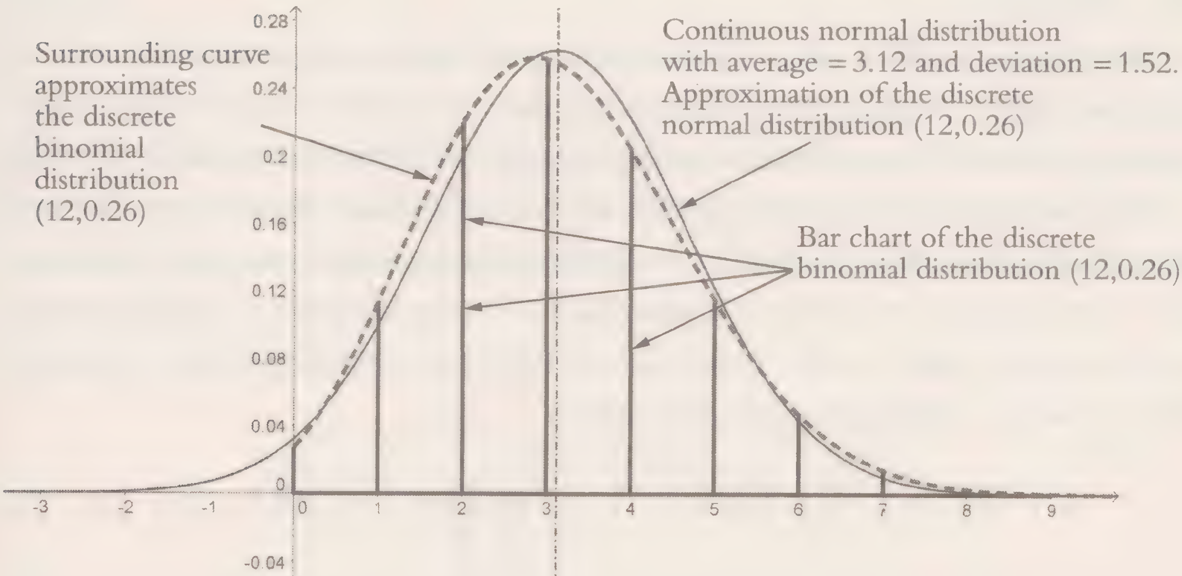


Diagram of probability bars PB(k) *in which* k *users of the brand appear in a sample of 12 households – if the use of the brand is estimated at being 26% of households. It is a discrete variable function with the Cartesian equation:*

$$\mathrm{PB(k)} = \binom{12}{\mathrm{k}} \cdot 0.26^{x} \cdot 0.74^{12-k}.$$

It is worth remembering that $\binom{12}{\mathrm{k}} = \frac{12!}{(12-\mathrm{k})!\mathrm{k}!}$.

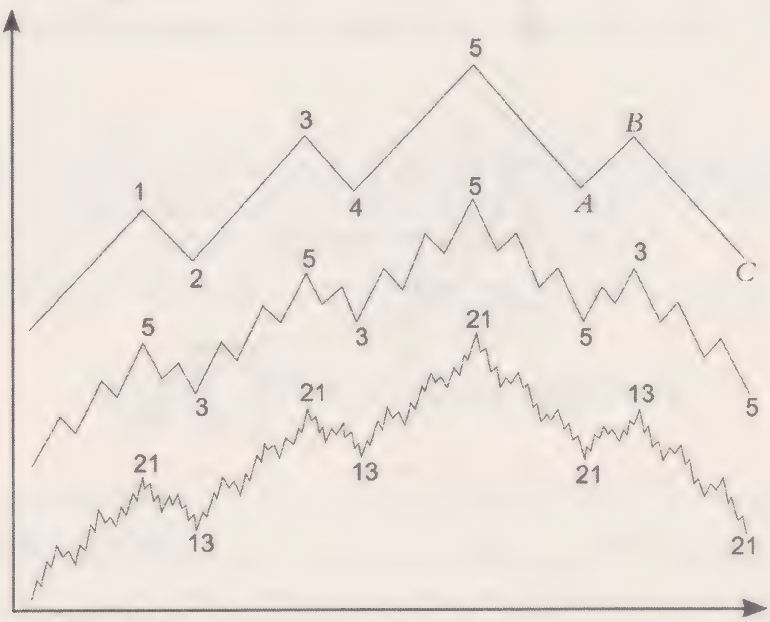
When the number of households surveyed is very high, the discrete binomial distribution is close to a normal distribution where the average is $12 \cdot 0.26 = 3.12$ and the standard deviation $\sigma = \sqrt{12 \cdot 0.26 \cdot 0.74} = 1.52$, which is a continuous exponential function from the Cartesian equation:

$$PN(x) = \frac{1}{1.52\sqrt{2\pi}} e^{-\frac{(x-3.12)^2}{2 \cdot 1.52^2}}.$$

Curves of the stock exchange

Ralph Nelson Elliott (1871-1948) made a considerable contribution to the behaviour of the stock exchange. Elliott waves explain the oscillations that the stock exchange experiences. The American economist elaborated a very long series of data from the New York Stock Exchange and studied the changes in the Dow Jones index. From the analysis of this empirical curve he concluded that the evolution exhibited certain repetitive patterns with different periods and amplitudes.

Elliott waves explain that stock markets have a rhythm that is repeated, following a periodic curve. Their mathematical equation is unknown, as it involves an empirical curve. However, their repetition and known form allows analysts to make certain predictions. The upward stretch has five ascending waves followed by another three descending waves. The complete cycle has eight waves, five going up and three going down. In the ascending part of the cycle, waves 1, 3 and 5 are impulsive, while 2 and 4 correct waves 1 and 3. After five ascending waves, a period of descent begins with three waves A, B and C.



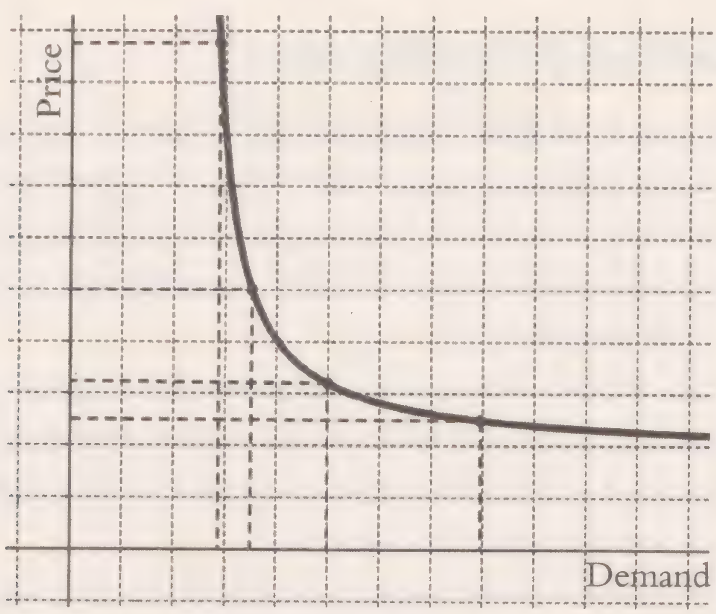
Elliott Waves (source: Masur). The equation is unknown as this involves an empirical curve.

Curves of the market

The market is as old as we are, but it is now a highly evolved and complex mechanism. It is possible to produce goods thousands of miles away and for these to pass through many intermediary transactions and sales. A small group of manufacturers (supply) may come together and agree to fix prices, whereas large groups of consumers may also exert pressure on prices. However, in spite of this, the market continues to be the mechanism whereby buyers (demand) and sellers (suppliers) meet to set prices. The equilibrium market price is the price at which buyers and sellers reach an agreement under the conditions called perfect competition, which is an ideal situation.

Demand is studied using historic statistical tables for the level of demand of a product, in line with the changes in prices. Each price corresponds to a level of demand.

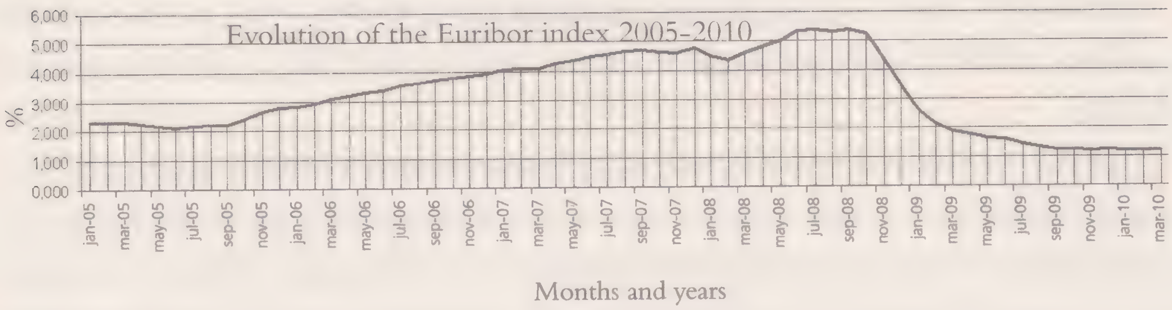
The price-demand curve follows a mathematical curve called a hyperbola, with prices represented on the vertical axis and the levels of demand on the horizontal axis. This curve is the graphical representation of inverse proportionality, a process that works as follows: the more there are of one of the variables, the fewer there are of the other). In this case, the lower the price of the goods in the market, the greater the demand for it.



Hyperbola curve. Equation in Cartesian coordinates: $y = \frac{1,000}{x}$.

A curve for mortgages

The Euribor index, which is used to calculate the interest applied to mortgages in Europe, varies every month. We can generate a curve that indicates its evolution during a period of time. However, this entails a curve elaborated empirically, in other words from data collected over a long period of time. This data is the compilation every month of this index. Therefore, it lacks a mathematical equation. However, the analysis of this empirical curve allows us to gather data about the historic evolution of world finances and the social events that have produced sudden changes in them.



Evolution of Euribor in the period 2005–2010. The equation or formula is unknown as it follows an empirical curve.

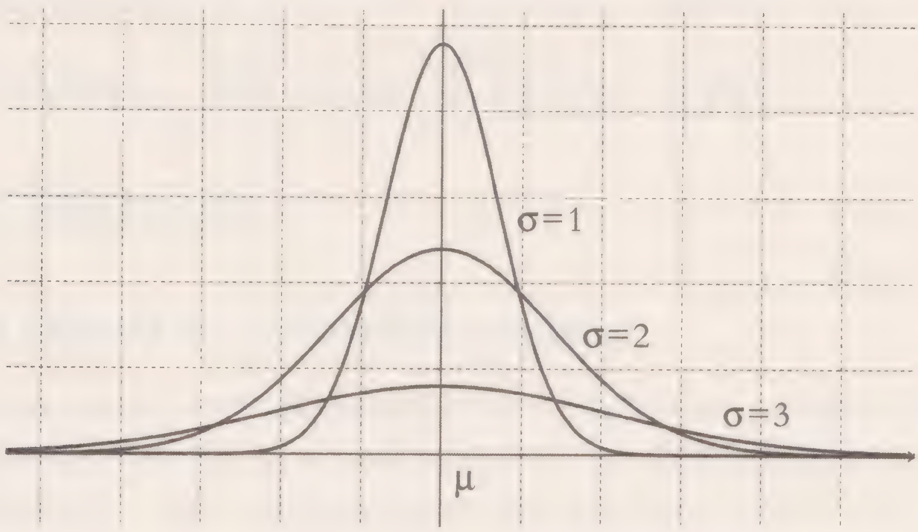
Normal or Gaussian probability curve

This curve is used a lot in all types of statistical applications. It was developed by Carl Friedrich Gauss at the beginning of the 19th century. It is used extensively because numerous phenomena tend to behave following the values of this equation, and many random variables (with random values) that may be continuous (can adopt values of all types) follow a bell-shaped curve.

The central value μ is the average value and the one that appears most frequently. The number σ (standard deviation) indicates the degree of variability of values with respect to the average. A distribution with a large σ indicates that there are more values far from the average value than close to it.

The normal probability curve is significant due to the fact that there are many variables associated with natural phenomena that follow this mathematical model. For example, certain anatomical characters of people, such as size or weight; physiological, such as the effect of a dose of a drug on different individuals; sociological, such as the consumption of a product by a group of people;

psychological, such as IQs; and others, such as the deviation of the required quality of a part manufactured by a machine in a factory.



The normal distribution or Gaussian bell curves with various standard deviations. The equation in Cartesian coordinates is

$$PN(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}.$$

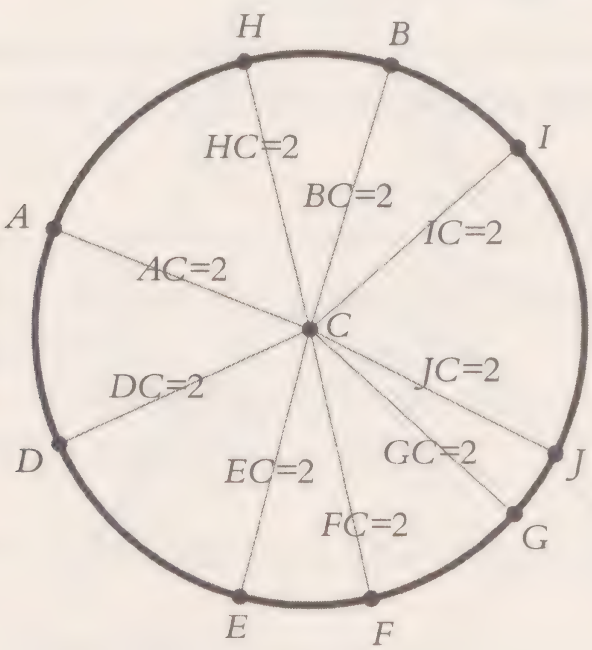
Chapter 2

Curves: How They are Drawn, How They are Measured

Curves defined by a geometric property

So far some curves have been defined by their equations using three types of reference systems: Cartesian, polar and parametric coordinates. However, curves can also be defined by a geometric property, without needing to know their equation. Each geometric property ensures the form of the corresponding curve. Examples of this other type of curve definition, the circle, the ellipse, parabola, tractrix and the lemniscate are described as follows.

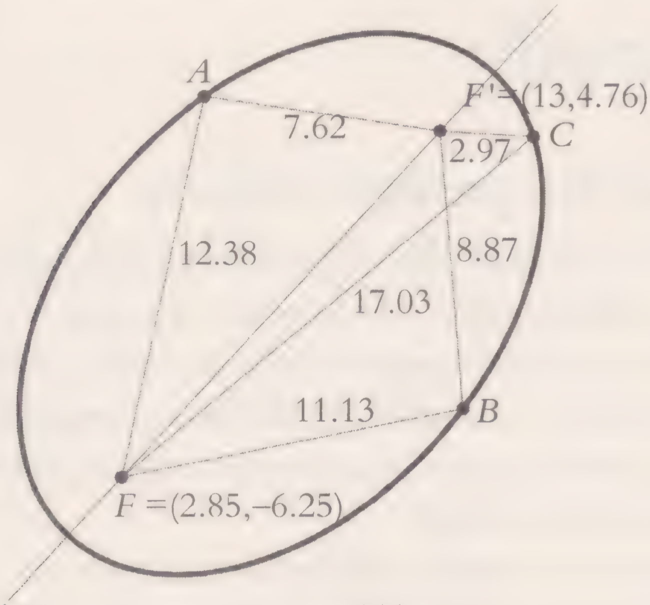
The *circle curve* with a radius of 2 and its centre on point C can be defined geometrically as the set of points (or the geometric position of the points) which occur at a distance of 2 from point C.



We can ascertain the geometric property that defines this circle, given that: $AC=BC=DC=EC=FC=GC=HC=IC=JC=2$.

If C is the point (–4,3), its Cartesian equation would be: $(x+4)^2+(y-3)^2=4$. All the points of an ellipse curve fulfil the geometric condition that the sum from any of the points to the two fixed points F and F' (focal points) always results in a constant quantity (the length of the greatest diameter of the ellipse $2a$). We can determine the geometric property that defines this ellipse, given that:

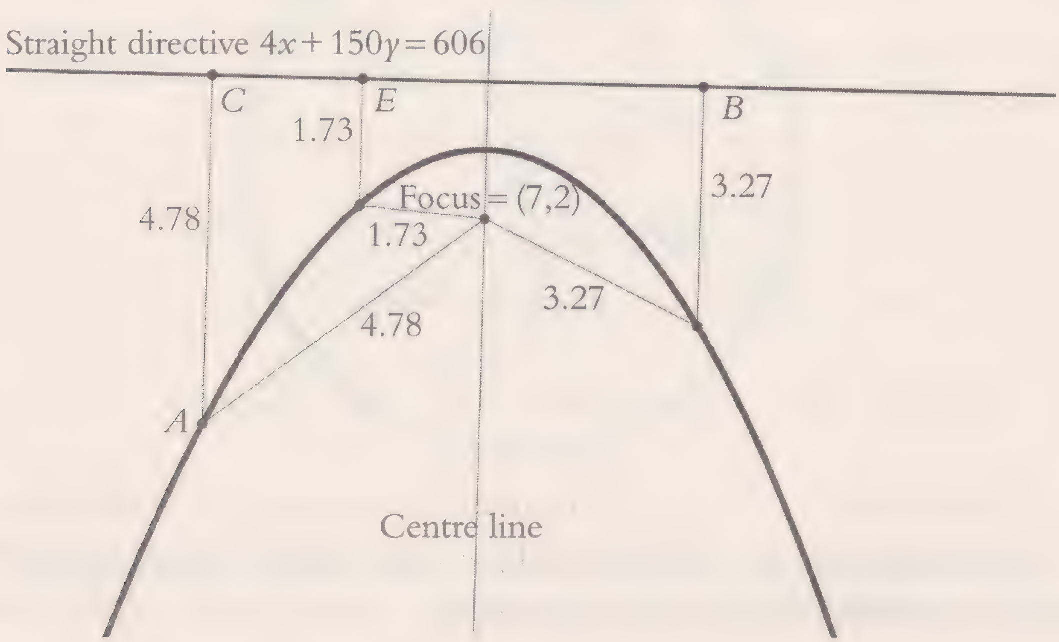
$$AF+AF'=BF+BF'=CF+CF'=20=2a.$$



And its complex Cartesian equation would be:

$$118747x^2-89498xy+111459y^2-1949383x+875856y+1032543=0.$$

A parabola curve is defined as the set of points located an equal distance from a point F (focal point of the parabola) and from a line (guideline of the parabola).



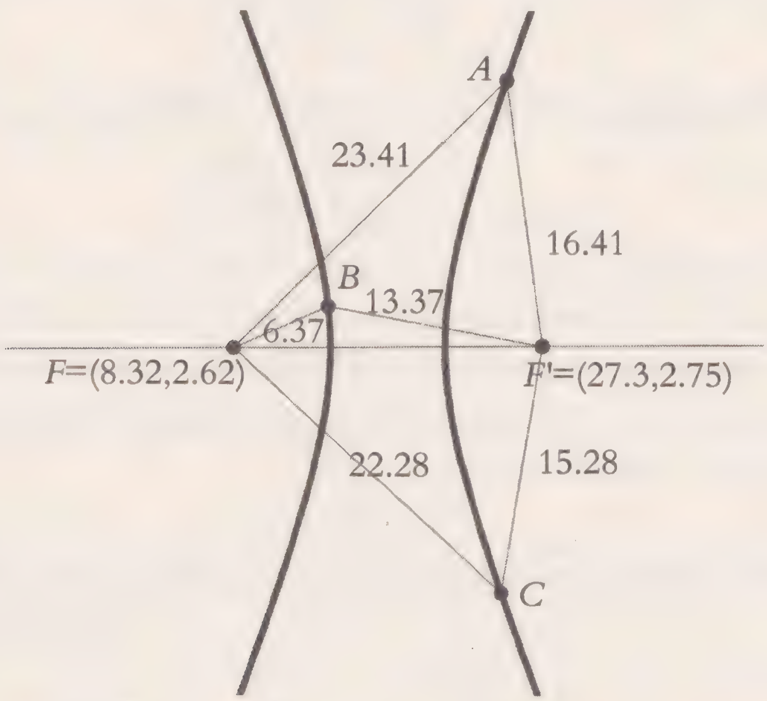
We can determine the geometric property that defines this parabola at its points A, D and P, given that:

$$AC = AF = 4{,}78;\ DE = DF = 1{,}73;\ PB = PF = 3{,}27.$$

And its complex Cartesian equation would be:

$$225x^2 - 11xy - 3103x + 919y + 8239 = 0.$$

The points of the hyperbola curve are defined by the geometric property that the difference between the distances of each one of them from the two fixed points (focal points of the hyperbola) is a constant value. In the hyperbola in the diagram this value is 7.



We can determine the geometric property that defines this hyperbola at its points A, B and C, given that:

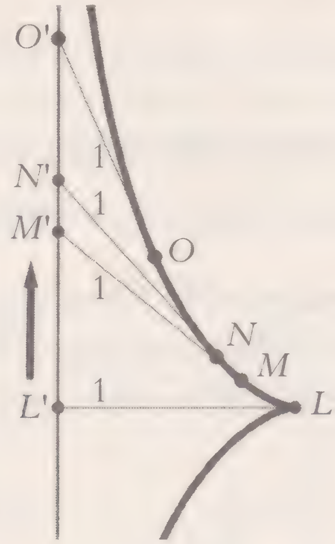
$$CF - CF' = BF - BF'' = AF - AF'' = 7.$$

And its complex Cartesian equation would be:

$$15.895\,x^2 + 159.994\,y^2 - 1.847xy - 61.206\,x - 826.525\,y + 4.518.289 = 0.$$

The tractrix curve is defined geometrically as the trajectory that follows a point (situated initially on L) when it is pulled by another (situated initially on L'), which maintains a constant distance (in this case the unit has been selected). L' moves

through a straight line (taking positions M', N', O'). The trajectory is defined by points such as M, N, O, etc.



The complex Cartesian equation of the arms of the *tractrix* is:

$$y=\pm(\sqrt{1-x^2}-\ln(\frac{1+\sqrt{1-x^2}}{x})).$$

The points of the lemniscate curve are defined by the geometric property that the product of the distances from each one of them to the two fixed points (focal points of the lemniscate) results in a constant value. In the lemniscate considered between –1 and +1, this value is 0.5.

Point	r_1	r_2	$r_1 \cdot r_2$
G	0.29	1.69	0.490
T	0.29	1.71	0.496
C	0.3	1.66	0.498
A	0.31	1.6	0.496
H	0.33	1.51	0.498
E	0.36	1.4	0.504
B	0.4	1.25	0.500
I	0.47	1.05	0.494

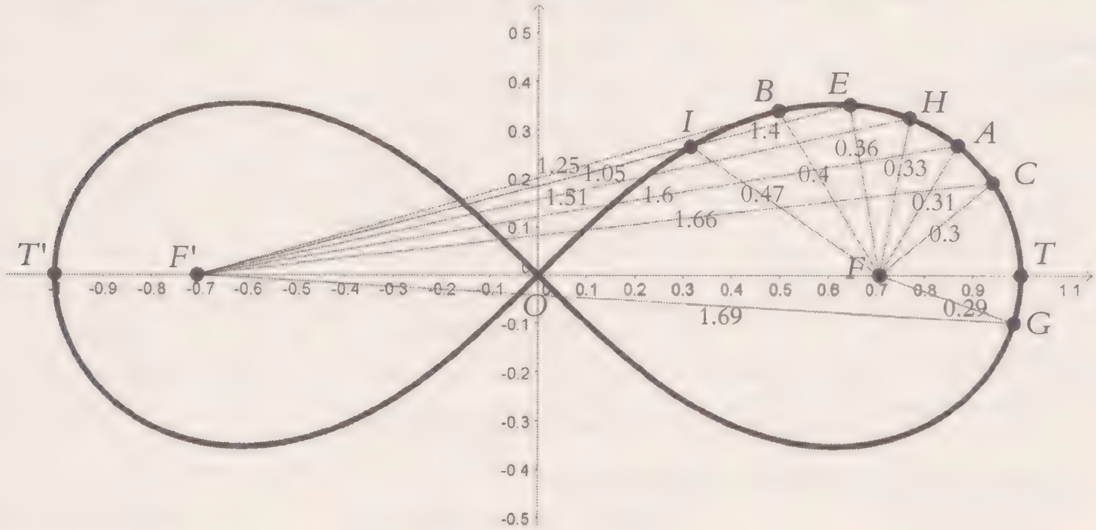
Approximating the lengths of the segments, the constant value sought gives values close to 0.5 for the selected points. These small differences are due to the low precision graphic representation software used. However applying the property of the extreme points F and F', we have:

$$OF \cdot OF' = OF^2 = (1-OF)\,(1+OF') = 1-OF^2.$$

As OF and OF' measure the same, isolating OF will give: $OF^2 = 1 - OF^2$, and hence:

$$2 \cdot OF^2 = 1;\ OF^2 = \frac{1}{2};\ OF = \frac{1}{\sqrt{2}} \cong 0.707.$$

In the figure an approximate value has been taken for the focal points $OF = 0.7$.



As $a = 1$, the equation of this lemniscate in Cartesian coordinates would be:

$$(x^2 + y^2)^2 = x^2 - y^2.$$

In contrast, the polar equation is much simpler:

$$r = a\sqrt{\cos 2\theta}.$$

Since its discovery by Jakob Bernoulli, the lemniscate curve has been adopted as the symbol for infinity.

Curves defined by a formula or equation

This representation method is the most common nowadays thanks to the numeric and symbolic calculation facilities of modern computing.

Some of the numerous types of curves that exist are combinations of others that are already known. They are depicted using the operations that are described in their equation. For example, the curve defined by the equation

$$y = \frac{x}{x^2 - 1}$$

can be drawn by dividing the ordinates of the curve $y = x$ (a line) by those of the curve $y = x^2 - 1$ (a parabola). An Excel spreadsheet can produce it using this data:

Point	x	y line $y=x$	y parabola $y=x^2-1$	$\frac{y \text{ straight line}}{y \text{ parabola}}$; $y=\frac{x}{x^2-1}$	Coordinates point
E_1	–2.5	–2.5	5.25	–0.48	(–2.5,–0.48)
D_1	–2	–2	3	–0.67	(–2,–0.67)
C_1	–1.5	–1.5	1.25	–1.20	(–1.5,–1.2)
B_1	–1	–1	0	$-\infty$	$(-1,-\infty)$
A_1	–0.5	–0.5	–0.75	0.67	(–0.5,0.67)
O	0	0	–1	0.00	0
A	0.5	0.5	–0.75	–0.67	(0.5,0.67)
B	1	1	0	$-\infty$	$(1,-\infty)$
C	1.5	1.5	1.25	1.20	(1.5,1.2)
D	2	2	3	0.67	(2,0.67)
E	2.5	2.5	5.25	0.48	(2.5,0.48)

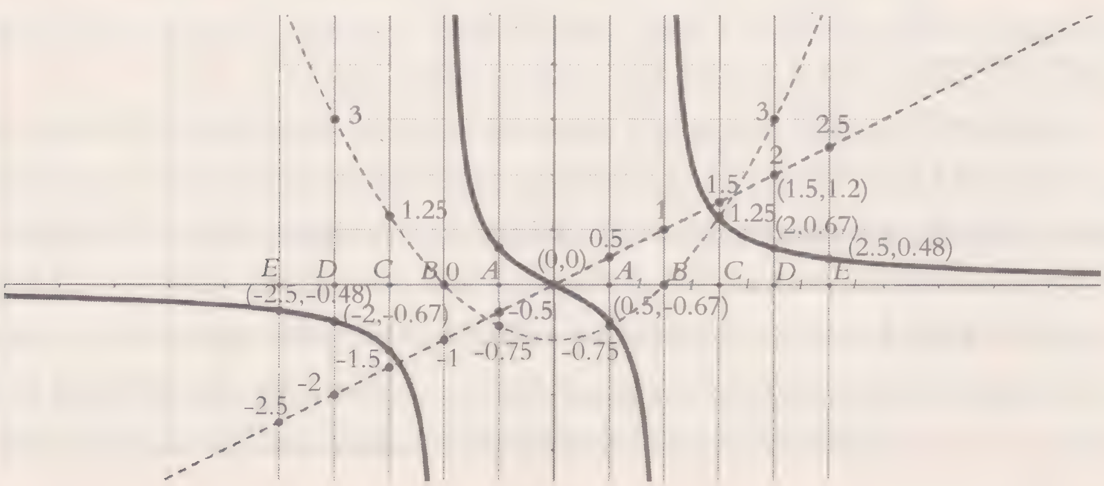
In the calculation of the curve's points

$$y=\frac{x}{x^2-1},$$

the values close to $x=-1$ and $x=+1$ lead to values of y that tend to be very large or very small numbers (according to whether they are positive or negative).

It is often commented that the points that have values close to x=–1 on the left side produce very small values of y (they say that y tends to $-\infty$) and the points that have values close to x=–1 on the right side produce very large numbers of y (tends to $+\infty$). With the data from the diagram and the table of values we can understand the resulting form of the rational curve:

$$y=\frac{x}{x^2-1}$$



To see the figure better, a different scale has been chosen for axis X and axis Y. These rational curves have an equation that is the quotient of another two polynomial curves of a variable. A polynomial of a variable is an addition or subtraction of expressions in which the powers of the variable appear with whole and positive exponents, multiplied by numeric constants. Some examples of polynomials are included to give us a closer look at the concept.

Expression	Type of polynomial	Terms
$3x^2-1$	Second-degree polynomial at x	It has two terms: $3x^2$, -1
x	First-degree polynomial at x	It has one term: x
1	Zero-degree polynomial at x	It has one term: 1
$(3x^2-1)^2=9x^4-6x^2+1$	Fourth-degree polynomial at x	It has three terms: $9x^2$, $-6x^2$, $+1$
$x\sqrt{2}-6x^2+3$	Second-degree polynomial at x	It has three terms: $x\sqrt{2}$, $6x^2$, 3
$\dfrac{3}{x^2-1}$	It is NOT a polynomial	—
$13\sqrt{x^3}+26x^2-18$	It is NOT a polynomial	—

A *polynomial curve* always has a Cartesian equation of the form $P(x,y)=0$, where $P(x,y)$ is a polynomial of the variables x and y of any level. An example is the curve: $225x^2-11xy-3.103x+919y+8.239=0$. The curve

$$y=\frac{x}{x^2-1},$$

which has been drawn by means of the quotient of the ordinates and the points on the graph of $y=x$ and $y=x^2-1$, is no longer of a polynomial nature.

The curves, with their equations being the quotient of two polynomials, are called *rational curves*. To extend the practice of drawing curves of this type, a similar process is applied to that of the previous curve to draw a new curve:

$$y=\frac{1}{(3x^2-1)^2}.$$

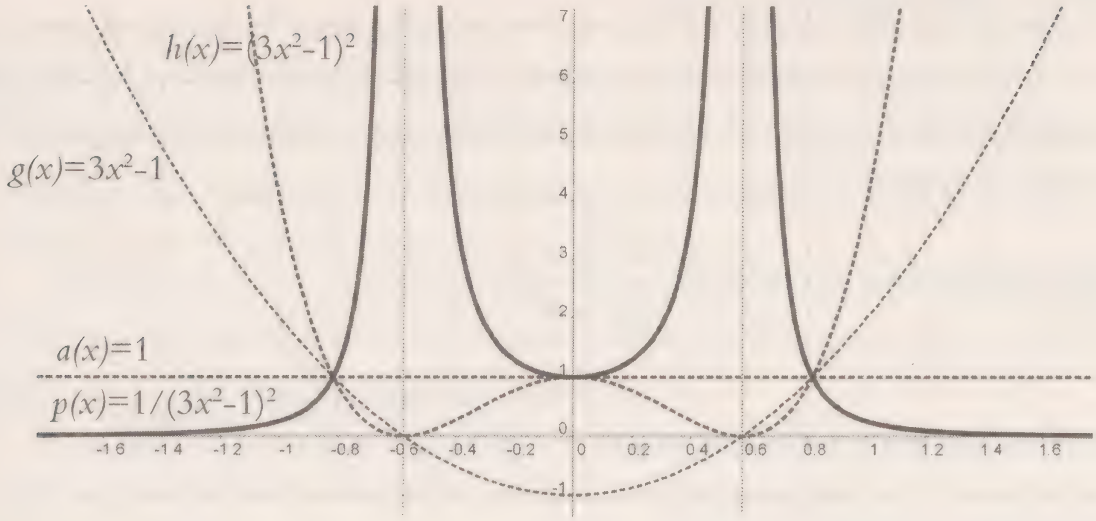
By analysing the table of values and the accompanying graph of the curves we can see the polynomial curves $y = 1$ and $y = (3x^2 - 1)^2$.

Abscissa x	Line $y = 1$	Quadratic $y = (3x^2 - 1)^2$	$\frac{y \text{ straight line}}{y \text{ parabola}}$, $y = \frac{1}{(3x^2 - 1)^2}$
–2.000	1.000	121.000	0.008
–1.750	1.000	67.035	0.015
–1.500	1.000	33.063	0.030
–1.250	1.000	13.598	0.074
–1.000	1.000	4.000	0.250
–0.750	1.000	0.473	2.116
$-\frac{1}{\sqrt{3}} = -0.577350269$	1.000	0.000	$+\infty$
–0.500	1.000	0.063	16.000
–0.250	1.000	0.660	1.515
0.000	1.000	1.000	1.000
0.250	1.000	0.660	1.515
0.500	1.250	0.910	1.765
$+\frac{1}{\sqrt{3}} = +0.577350269$	1.000	0.000	$+\infty$
0.750	1.000	0.473	2.116
1.000	1.000	4.000	0.250
1.250	1.000	13.598	0.074
1,500	1.000	33.063	0.030
1.750	1.000	67.035	0.015
2.000	1.000	121.000	0.008

By dividing

$$\frac{1}{(3x^2 - 1)^2}$$

we can approximately obtain the indicated points that form of the curves in the following figure:



This shows that the points of a rational curve in which the value of y soars (tends to $+\infty$ or to $-\infty$) correspond to the abscissa x which give a result of zero in the polynomial denominator of the equation of the curve. In the case where

$$y = \frac{x}{x^2 - 1},$$

the curve soars to the values of x which make the denominator $x^2 - 1 = 0$, or, $x^2 = 1$; $x = +1$, $x = -1$. In the case where

$$y = \frac{1}{(3x^2 - 1)^2},$$

the curve soars to the values of x which make the denominator $(3x^2 - 1)^2 = 0$, or, $3x^2 = 1$; $x = +\frac{1}{\sqrt{3}}, x = -\frac{1}{\sqrt{3}}$.

The vertical lines in which the function soars (y tends to be towards $+\infty$ or $-\infty$) are called vertical asymptotes of the rational curve. In the curve

$$y = \frac{x}{x^2 - 1}$$

the asymptotes are vertical lines: $x = +1$ and $x = -1$. In the curve

$$y = \frac{1}{(3x^2 - 1)^2}$$

the asymptotes are the: $x = +\frac{1}{\sqrt{3}}$ and $x = -\frac{1}{\sqrt{3}}$ (in the previous graph they approximated $+0.6$ and -0.6).

Explicit and implicit curves

Some curves have an equation the variable of which is y (dependent variable) is completely isolated in one of its two sides. These curves are said to be in *explicit* form, such as the following:

The rational type of curve: $y=\frac{1}{(3x^2-1)^2}$;

the tractrix in Cartesian coordinates: $y=\pm(\sqrt{1-x^2}-\ln(\frac{1+\sqrt{1-x^2}}{x}))$;

the circle: $y=\pm\sqrt{25-(x-2)^2}$;

the ellipse $y=\pm 5\sqrt{(1-\frac{x^2}{49})}$.

In other cases, the two variables are found intermingled within the equation and it is almost impossible to completely isolate one of them. In that case, the curves are in an *implicit* form, such as for the lemniscate $(x^2+y^2)^2=a(x^2-y^2)$; the circle $(x-2)^2+y^2=25$, or the ellipse

$$\frac{x^2}{49}+\frac{y^2}{25}=1.$$

However, in general, all the curves with a known equation can be drawn without great difficulty using suitable curve representation software, such as Geogebra or Derive.

Curves and functions

Curves are closely related to functions. In Euler's era (the 18th century; he was one of the best mathematicians who has ever lived), there was practically no difference between one concept and the other. At that time, a function was a curve where you could trace its entire length with a pencil without ever having to lift it from the paper and without it crossing itself on its path; it was said to be 'continuous'. Furthermore, it was required that the curve of a function did not have points 'at an angle' that would destroy the 'smoothness' of its path.

Since Euler there have been many changes in the concepts of curves and functions, but this intuitive vision from that era still continues to be adequate

for the majority of mathematical curves used in everyday life and in the professional field.

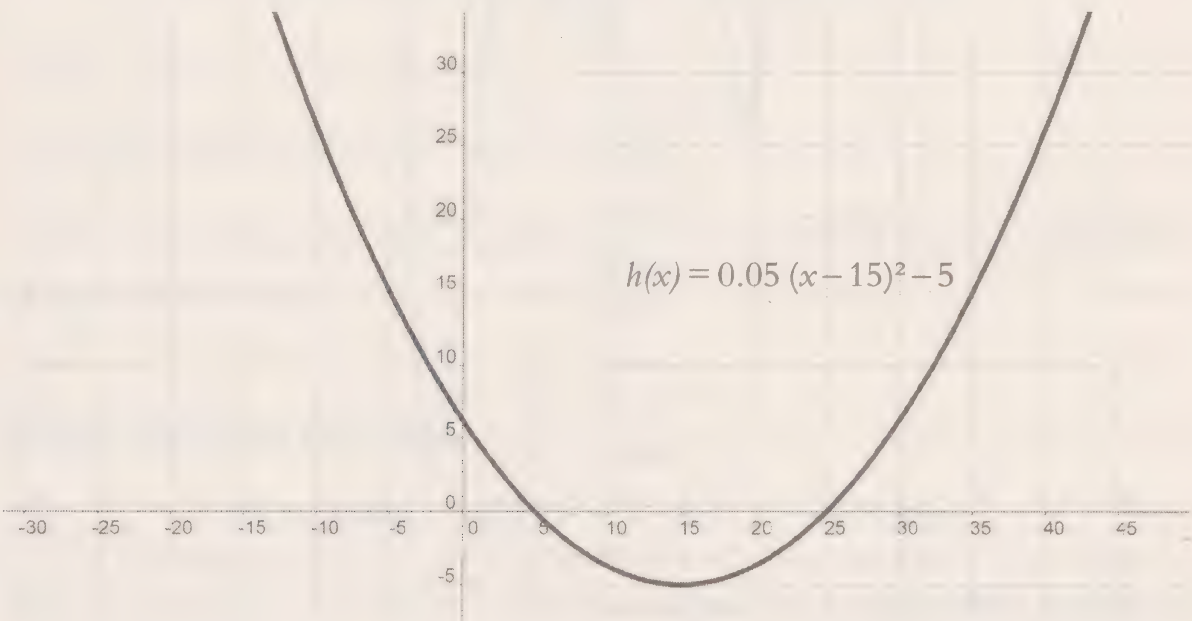
A function is the entire connection between the values of an independent variable x and another y dependent on it, and as such each value of x corresponds to a single value of y. In this case it is said that y is a function of x.

Well-known curves such as the circle, ellipse, parabola from a horizontal symmetry axis, lemniscate and tractrix are not strictly functions. We can see in their graphs that each value of x corresponds to two values of y.

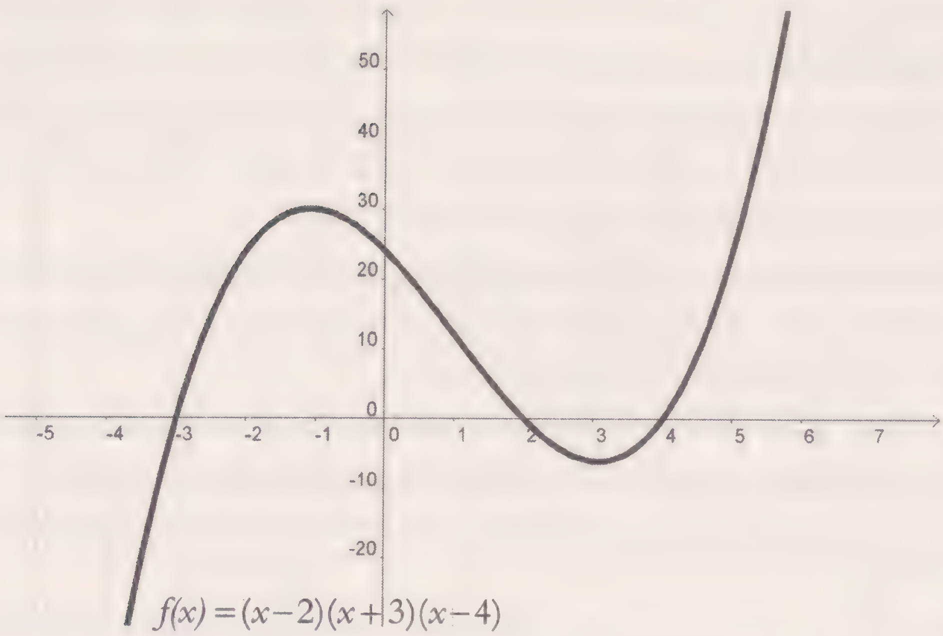
Rational curves are functions that have some exceptions in their continuous form. They are continuous except in the values of x in which the curve 'soars' to infinity. In $y=\dfrac{x}{x^2-1}$ the curve is continuous for all the values of x except for $x=+1, x=-1$.

In $y=\dfrac{1}{(3x^2-1)^2}$ the curve is continuous for all the values of x except for $x=+\dfrac{1}{\sqrt{3}}, x=-\dfrac{1}{\sqrt{3}}$.

In contrast, all the polynomial functions are always continuous, such as the second, third and sixth-degree functions highlighted below. And, evidently, all straight lines are first-degree polynomial functions. Polynomial functions correspond to known polynomial curves, such as the second-degree parabola $h(x)=0.05(x-15)^2-5$, a second-degree polynomial curve:



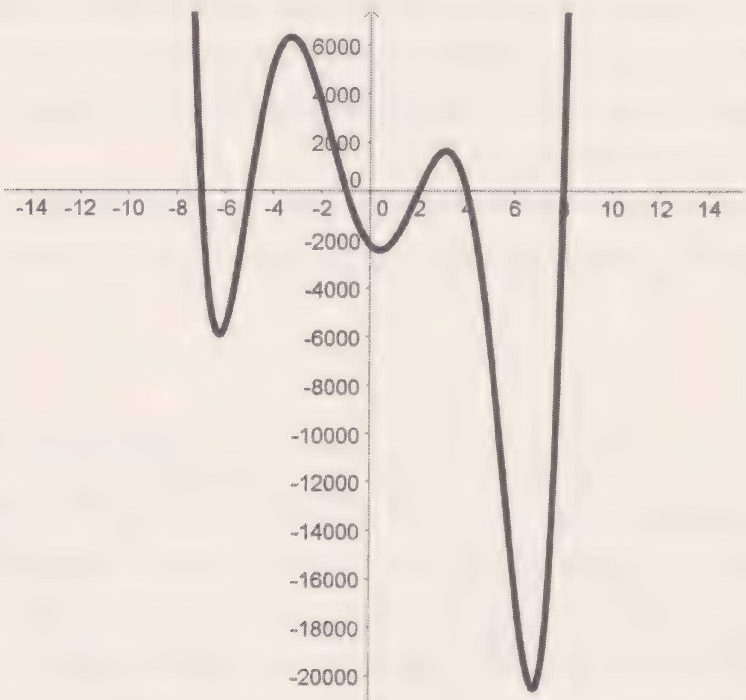
Or the third-degree polynomial function: $y=x^3-3x^2-10x+24$.



$f(x) = (x-2)(x+3)(x-4)$

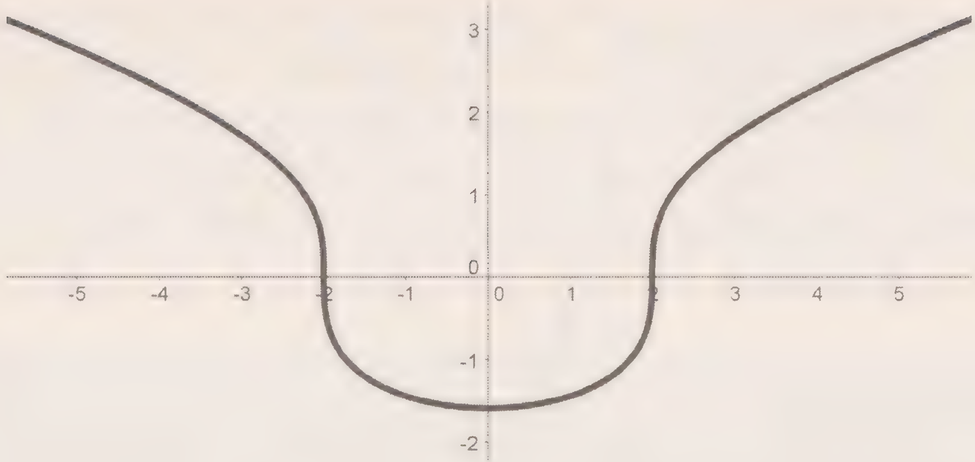
Or the sixth-degree polynomial function:

$$y = x^6 - x^5 - 79x^4 + 41x^3 + 1310x^2 - 1048x - 2240.$$

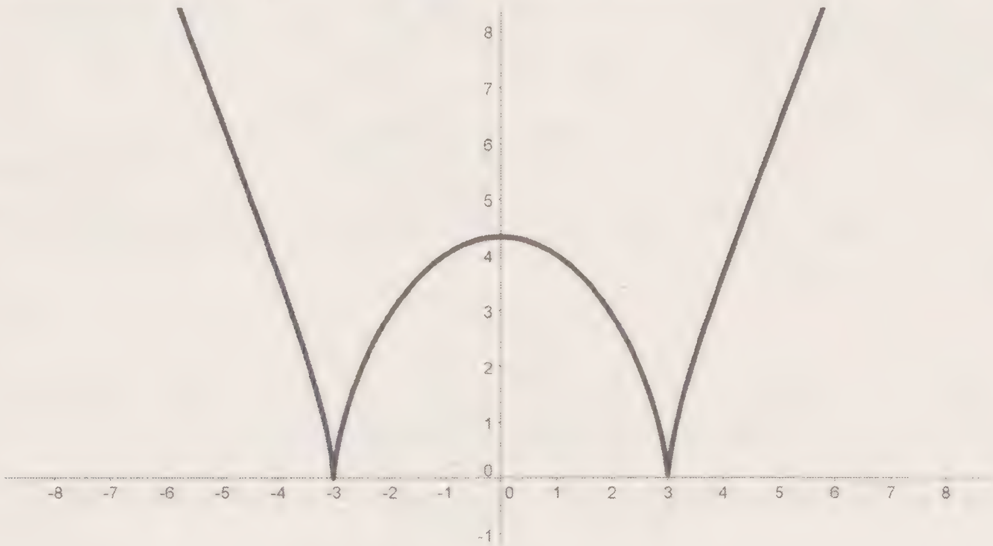


$g(x) = x^6 - x^5 - 79x^4 + 41x^3 + 1310x^2 - 1048x - 2240$

Rational functions correspond to the rational curves described previously. Irrational functions are those whose equation contains a variable under the root sign, such as the function $y = \sqrt[3]{x^2 - 4}$, which has the following graph:



Or the function $y=\sqrt[3]{(x^2-9)^2}$, the curve of which is:



Or the more complex curve $\sqrt{y^2-x^2}+\sqrt[4]{x^3+4}-xy+12=0$.

The set of polynomial functions, rational and irrational, are grouped into the set of algebraic functions.

Transcendental functions

The functions that cannot be expressed in the form of polynomials or roots are known as transcendental functions. The simplest are the exponential function, the logarithmic function and trigonometric and hyperbolic functions. The exponential function has a Cartesian equation of the type $y=k\cdot a^x$, with k being a real number, and a, a real positive number, such as, for example, $y=2\cdot 3^x$, $y=10^x$, $y=13\cdot(\sqrt{5})^x$.

Among the many exponential functions there exists a very particular one which is based on the constant number e, $y = e^x$. This *exponential curve* is unusual in that the gradient of the right tangent at each point is equal to the ordinate of the point, as can be seen in the following figure. This means that it is the only function whose derivative coincides with the same function, or in other words, that $y' = y = e^x$.

THE NUMBER *e*

The constant e was discovered by the Swiss scientist Jakob Bernoulli on studying the following problem of compound interest: If £1 is invested with an annual interest of 100% and the interest is paid once per year, at the end of the period the amount will be £2. If the interest is paid 2 times a year, dividing the interest by 2 (50% biannually), the amount obtained is £1 multiplied twice by 1.5, in other words:

$$£1 \cdot \left(1+\frac{1}{2}\right)^2 = £1 \cdot 1.5^2 = £2.25.$$

If the interest is paid in quarterly periods and we divide the annual interest by 4 (25% quarterly), at the end of the year this will be obtained:

$$£1 \cdot \left(1+\frac{1}{4}\right)^4 = £1 \cdot 1.25^4 = £2.4414...$$

In the case of monthly payments the capital accumulated will rise to:

$$£1 \cdot \left(1+\frac{1}{12}\right)^{12} = £2.61303... .$$

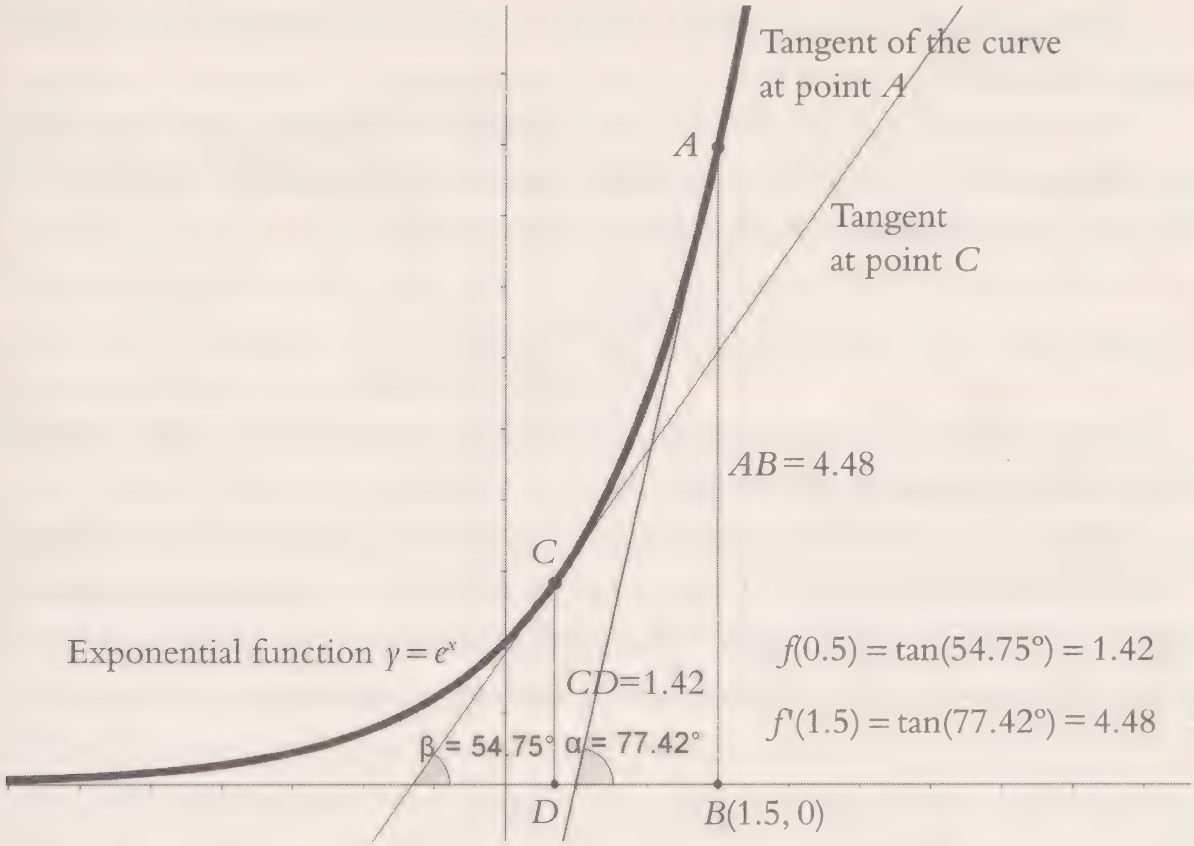
On increasing the number of payment periods n times (which tends to grow indefinitely) and dividing by n the interest rate in the period, the total amount obtained by £1 at the end of the year would be:

$$\lim_{n\to\infty}\left(1+\frac{1}{n}\right)^n.$$

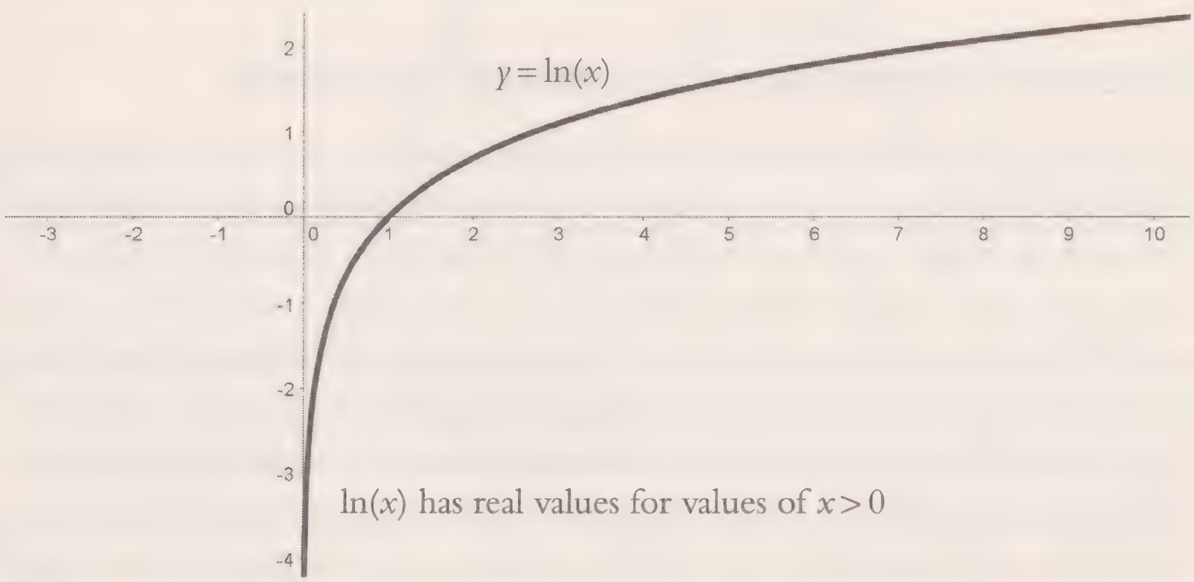
By giving increasing values to n, Bernoulli ascertained that this expression approximates to:

$$e = 2.7182818...$$

Hence in financial mathematics the number e is defined as the capital limit accumulated through an investment of £1 with an annual compound interest rate of 100% with continuous payment of interest. The first mathematician who gave this constant the name e was Leonhard Euler in 1727.



A logarithmic function is the inverse function of the exponential function. It results from to the question: "What exponent should a real number be raised to (called the logarithmic base) to obtain another determined number?" The expression: $\log_a N = x$ means that the number a (the base) elevated to x should give N, or rather: $a^x = N$. The *l*ogarithmic curve based on e, with equation $y = \ln(x)$, looks as follows:



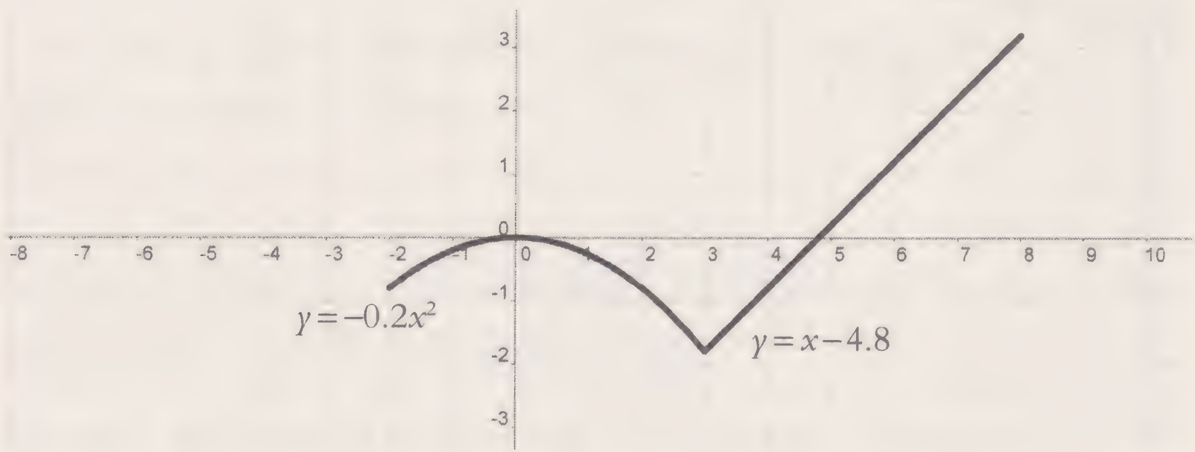
It only exists for positive values of x , and in $x=0$ the y 'soars' towards $-\infty$ (with a vertical asymptote).

Other types of functions are those functions defined in sections, those that present a different equation according to the set of values of the independent variable x. An example is the following function defined in two sections:

$$y(x)=\begin{cases}-0.2x^2 \text{ for values } -2\leq x<3\\ x-4.8 \text{ for values } 3\leq x\leq 8\end{cases}$$

It is a parabolic curve between the values of $x=-2$ and $x=3$, and with a straight line between the values of $x=3$ and $x=8$.

Point (3,–1.8) is a 'strange' point where the function has two different tangent lines depending on whether the first or second equation is considered, given that it seems the point is located in both sections at the same time, as is shown in the figure. The mathematics of these types of functions can get complicated.



Gradients, tangent lines of a curve and derivatives

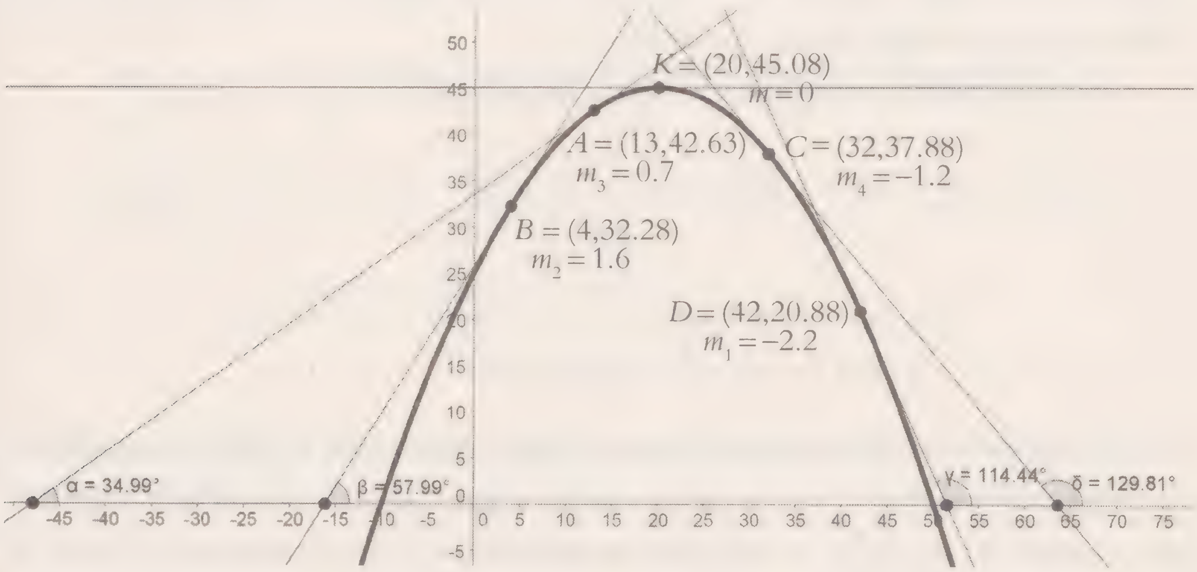
To find out the degree of variation or trend of the dependent variable y with respect to the independent x at point A of the curve, we are required to know if the function that represents them is increasing by a lot or by a little at point A. This allows us to compare its level of variation with respect to the value at other points on the curve. We use the gradient of the point to measure this level of variation, calculating the slope of the straight line tangent to the curve at point A. This is determined by calculating the angle at which the straight line tangent at A meets the horizontal axis. If the angle at α is very high, this means that the curve has a steep gradient at this point – and vice versa. And, as had already been shown, the gradient m of each

one of the straight lines (in their Cartesian equations, $y = m \cdot x + n$) is the tangent of its angle of inclination. For example, at points B, A and K, the gradient is decreasing, and at points K, C and D, it is increasing.

From this concept, in the second half of the 17th century Isaac Newton and Gottfried Leibniz, independently designed a method of calculating these gradients at each point in the curve from its equation. Thereby we can now find out if the curve increases or decreases – and whether this is by a lot or a little. This mathematical method is known as *derivation of a function.*

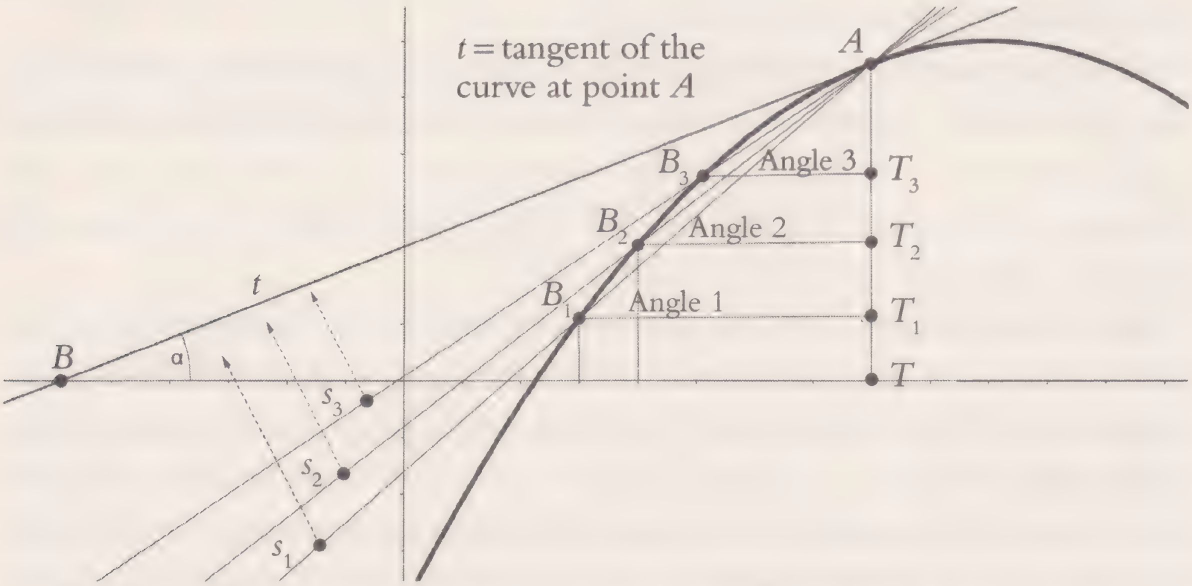
Both mathematicians also designed another method of determining, before plotting the curve on a graph, what points of the same variable y would acquire maximum and minimum values. This second method is called calculating the *extremes of a function.*

To tackle the first problem they started with the studies developed by Apollonius of Pergum (3rd century BC) and for the second, with the works of Fermat (beginning of the 17th century). To solve the so-called 'tangent problem', the pair doubtlessly drew, with utmost precision, a simple curve, such as the parabola in the following figure ($y = -0.05x^2 + 2x + 25.08$), and drew the different tangent lines of the curve, measuring their inclination angles and gradients.



Leibniz and Newton then devised a process for finding the gradient of point A by the so-called 'approaching the limit' method, which Newton called 'fluxional calculus'. They started with the gradients of the secant straight lines to the curve that pass through A, and the straight lines s_1, s_2, s_3,... that pass through the points of the curve, B_1, B_2, B_3. These secant lines approach point A little by little. Therefore,

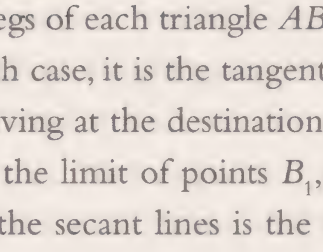
they tend to convert into the straight line t, which is precisely the tangent of the curve at point A:



The three small right-angled triangles AB_1T_1, AB_2T_2, AB_3T_3 decrease in size during this movement of the points B_1, B_2, B_3… towards A. When it arrives at A, the corresponding triangle 'disappears' (it has the size of zero).

The angles made by the secant lines s_1, s_2, s_3 with the horizontal axis decrease until the points B_1, B_2, B_3 coincide with A; then the angle with the horizontal axis is the same as the tangent line (α, alpha).

A right-angled triangle ABC complies in that: $\tan(A) = \frac{c}{b}$.



The quotient of the two legs of each triangle AB_1T_1, AB_2T_2, AB_3T_3 varies with the change of position. In each case, it is the tangent of angle 1, angle 2, angle 3,... up to points B_1, B_2, B_3,… arriving at the destination of their 'movement': point A. Mathematically it is said that the limit of points B_1, B_2, B_3,… is point A, and that the limit of the gradients of the secant lines is the tangent of the angle α of the tangent line t:

$$\frac{AT_1}{B_1T_1} = \tan\text{ (angle 1) } \frac{AT_2}{B_2T_2} = \tan\text{ (angle 2) } \frac{AT_3}{B_3T_3} = \tan\text{ (angle 3),... } \frac{AT}{BT} = \tan(\alpha)$$

The gradient of the line t (tangent line of the function $f(x)$ at point A) is

$$\lim_{B_i \to A} \tan(\text{angle } i) = \frac{AT}{BT} = \tan(\alpha).$$

The value of it

$$\lim_{B_i \to A} \tan(\text{angle } i) = \frac{AT}{BT}$$

is currently referred to as the derivative of the function $f(x)$ at point A and coincides with the value of the gradient of the tangent line of the function at point A.

In the case of the function under consideration $y = -0.05x^2 + 2x + 25.08$, Newton and Leibniz no doubt made a similar calculation to the following: Δx is the length in which the abscissa x varies between B_3 and A, and the length of the ordinate Δy y is between B_3 and A, therefore:

$$\frac{\Delta y}{\Delta x} = \frac{AT_3}{B_3T_3} = \tan(\text{angle } 3).$$

The equation of this curve can be expressed as $y = -0.05x^2 + 2x + 25.08$:

$$\frac{\Delta y}{\Delta x} = \frac{(-0.05(x+\Delta x)^2 + 2(x+\Delta x) + 25.08) - (-0.05x^2 + 2x + 25.08)}{\Delta x} =$$

$$= \frac{-0.05x^2 - 0.05 \cdot 2x \cdot \Delta x - \Delta x^2 + 2x + 2\Delta x + 25.08 + 0.05x^2 - 2x - 25.08)}{\Delta x} =$$

$$= \frac{-0.05 \cdot 2x \cdot \Delta x - \Delta x^2 + 2\Delta x}{\Delta x} = -0.05 \cdot 2x - \Delta x + 2.$$

When the length of Δx is decreasing and becoming very, very small, in other words when $\Delta x \to 0$, the value of the gradient at point A shall be the limit value:

$$\underset{\Delta x \to 0}{\text{limit}} \frac{\Delta y}{\Delta x} = \tan(\alpha) = \text{Derivative } f(x) \text{ in } A(x_0, y_0) = f'(x_0) = -0.05 \cdot 2 \cdot x_0 + 2.$$

In this case, point A is (13,42.63) and, therefore, $f'(13) = -0.05 \cdot 2 \cdot 13 + 2 = 0.7$, which coincides with the result of the gradient measured on the curve.

In the following table we can check at five points of the curve that the derivative of each point has the same value as the gradient at the given point.

Point	x	Function $f(x)$ $y=-0.05x^2+2x+25.08$	Derivative $f'(x)$ $y'(x)=-0.05\cdot 2x+2$	Gradient measured (m)
D	42	20.88	−2.2	tan(129.81°) = −2.2
C	32	37.88	−1.2	tan(114.44°) = −1.2
K	20	45.08	0	tan(0) = 0
A	13	42.63	0.7	tan(34.99°) = 0.7
B	4	32.28	1.6	tan(57.99°) = 1.6

The derivative function $y=-0.1x+2$ is used to calculate the gradient of the tangent at any point of the curve $y=-0.05x^2+2x+25.08$. In the case of functions of the type $y=k\cdot x^n$, its derivative shall be $y'=k\cdot n\cdot x^{n-1}$. For example, in the function $y=5x^4$, its derivative is $y'=20x^3$.

Applying the same method of approaching the limit used in the function $y=-0.05x^2+2x+25.08$ works for all kinds of functions. By experimenting with different functions the basic rules of derivation have been designed:

a) The derivative of a sum of various functions is the sum of its derivatives.
b) The derivative of the product of a constant for a function is the constant for the derivative of this function.
c) The derivative of a constant is zero.

For this reason, obtaining the functional derivative of a polynomial function is simple. It is sufficient to calculate the derivative of each one of its terms. In the case of the polynomial function $y=-3x^2+7x-2$, it is easy to check that its functional derivative would be: $y'=-6x+7$.

If we wish to calculate the derivative of any other type of function it is not necessary to repeat the process of approaching the limit used previously. Some rules of derivation are used for this that have been deduced by means of calculating the aforementioned limits for each type of function.

The tables of derivative functions of each type of function are used in the study of the tendencies of a function, or rather their areas and degrees of growth and reduction of a corresponding curve. The reader interested in studying this topic of derivation in more depth can refer to the works of analysis contained in the bibliography.

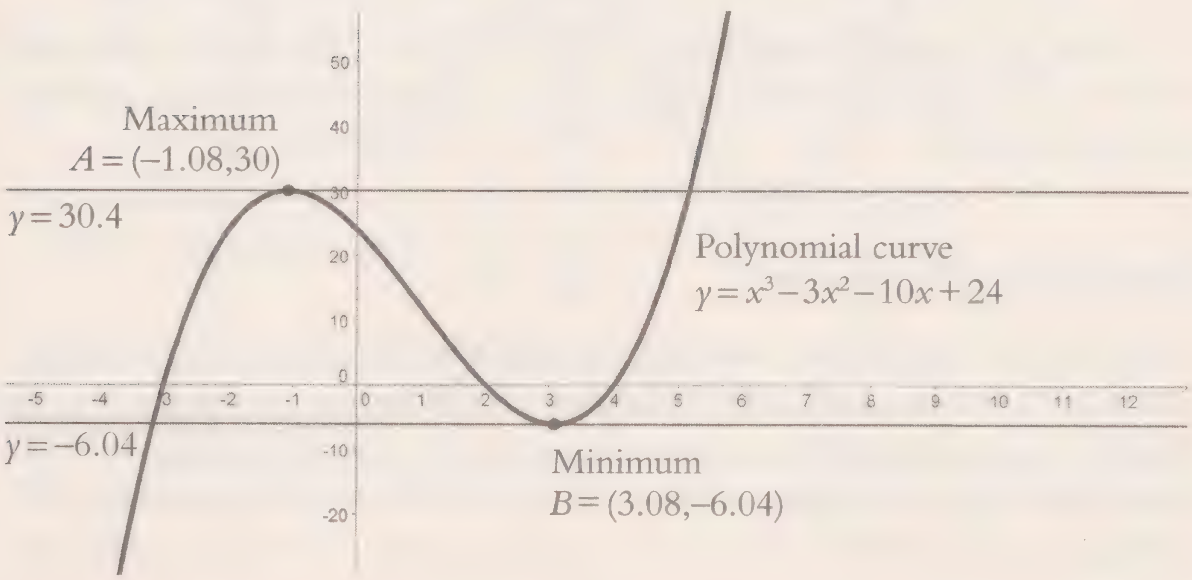
Extremes of a function

The second important problem tackled by Newton and Leibniz to determine the graph form of a function (its curve) was the determination of its extreme points, where the curve reaches its maximum or minimum value. This problem was tackled by Newton and Leibniz in the 1660s from the results of the analytical geometry developed by Pierre de Fermat: they applied the previously resolved problem of the tangents to the individual points of the curve in which the tangent line is a horizontal straight line.

Given that a horizontal line forms an angle of 0° with the horizontal and that the tangent of an angle of zero degrees has a value of zero, they managed to reduce the problem of determining the extremes of a function to the search for the values of x. This ensured that the first derivative (which is the gradient of the tangent line at this point) had a value of zero, or rather, to resolve the equation $f'(x)=0$. In the case of third-degree polynomial function $y=x^3-3x^2-10x+24$, its derivative is $y'=3x^2-6x-10$, and the values that give it a value of zero are the solutions to the second-degree equation $0=3x^2-6x-10$ which are obtained with the usual formula:

$$x=\frac{-b\pm\sqrt{b^2-4ac}}{2a}=\frac{6\pm\sqrt{(-6)^2-4\cdot 3\cdot(-10)}}{2\cdot 3}=\frac{6\pm\sqrt{36+120}}{6}=$$

$$=\begin{cases}\dfrac{6+\sqrt{156}}{6}=3.08\\[2ex]\dfrac{6-\sqrt{156}}{6}=-1.08\end{cases}$$



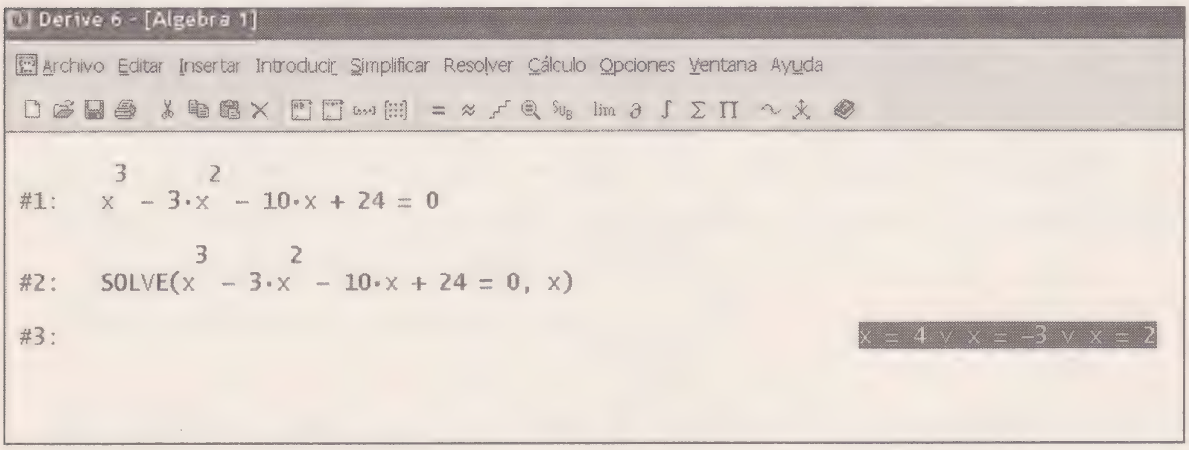
Zeros of a function

The zeros of a function $f(x)$ are the values of x in which the function is void, in other words, the values of x that fulfil: $f(x) = 0$.

Stated another way, it is the different points at which the curve cuts axis X, given that the points of the axis of abscissas are the only points on the Cartesian plane whose ordinates y are zero.

These intersecting points of the function with axis X are of great interest in the construction of the function graph, as they are the points at which the curve crosses axis X. To calculate these points the following equation needs to be resolved: $f(x) = 0$. Naturally, the difficulty of resolving this depends on the complexity of the equations of the function $y = f(x)$.

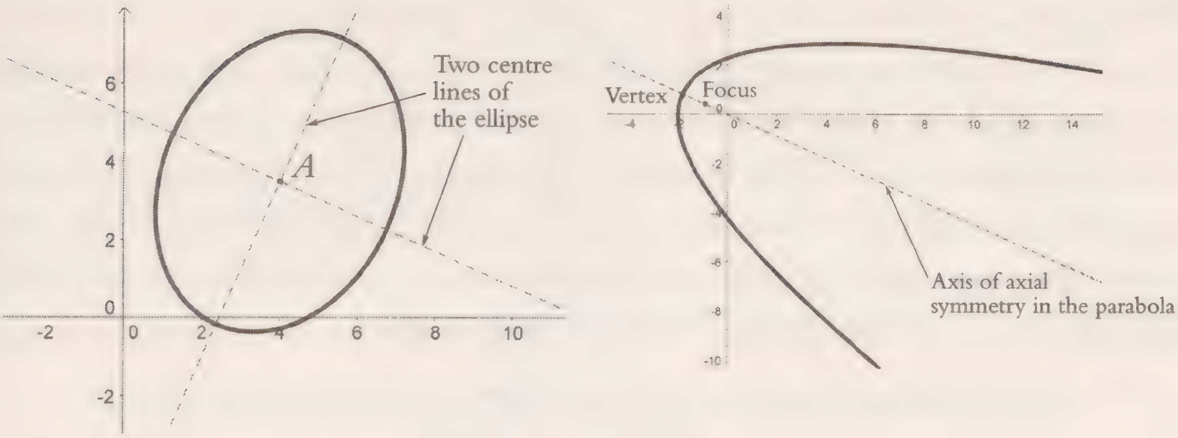
In the case of the curve $y = x^3 - 3x - 10x + 24$ the equation $0 = x^3 - 3x - 10x + 24$ needs to be resolved; this is a third-degree equation. For that we can use a suitable symbolic calculus software, such as Derive:



In this way we get three solutions: $x = 4$, $x = -3$, $x = 2$. Therefore, the curve cuts axis X at points (4,0) (–3,0) and (2,0), as can be seen in the graph of the curve on the previous page.

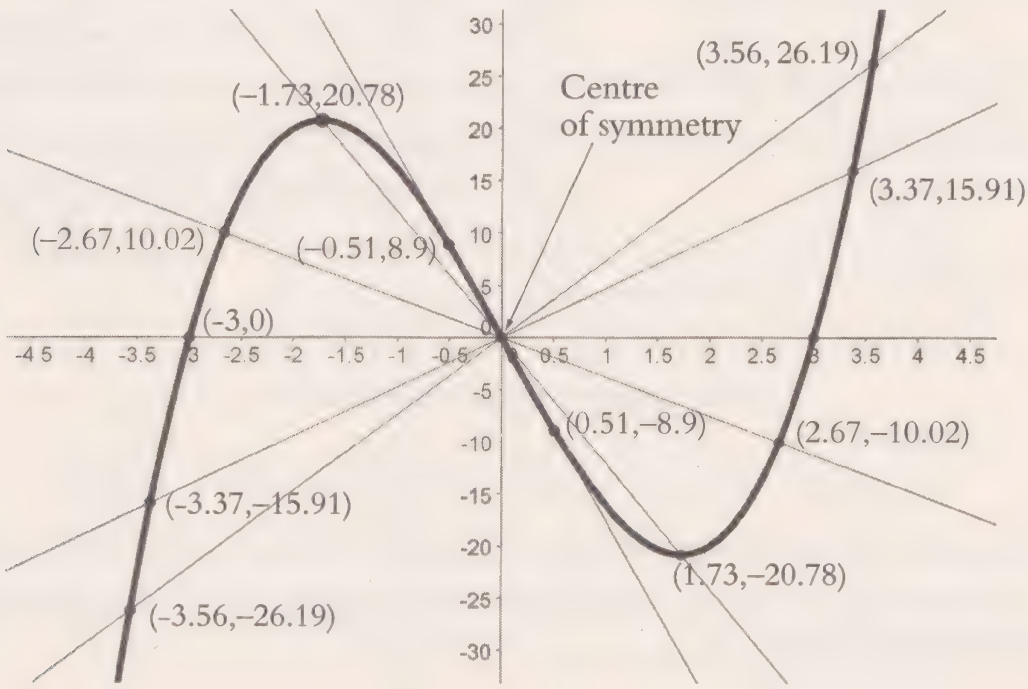
Symmetries of a function

The curve of some functions looks as if one part of it is its reflected image. This type of 'mirror' is what is called an axis of symmetry of the function. In this case it is said that the function has axial symmetry (symmetry with respect to an axis):



In the figures we can see the two axes of symmetry of the ellipse and its centres, as well as the single axis of symmetry of the parabola, its focal point and its vertex.

Another type of symmetry exists in curves, *central symmetry*. In this, all the points of the curve are symmetric with respect to a point called the centre of symmetry (or rather, the symmetric points are located at an equal distance from the centre). The following figure demonstrates a polynomial function of third degree with a centre of symmetry.



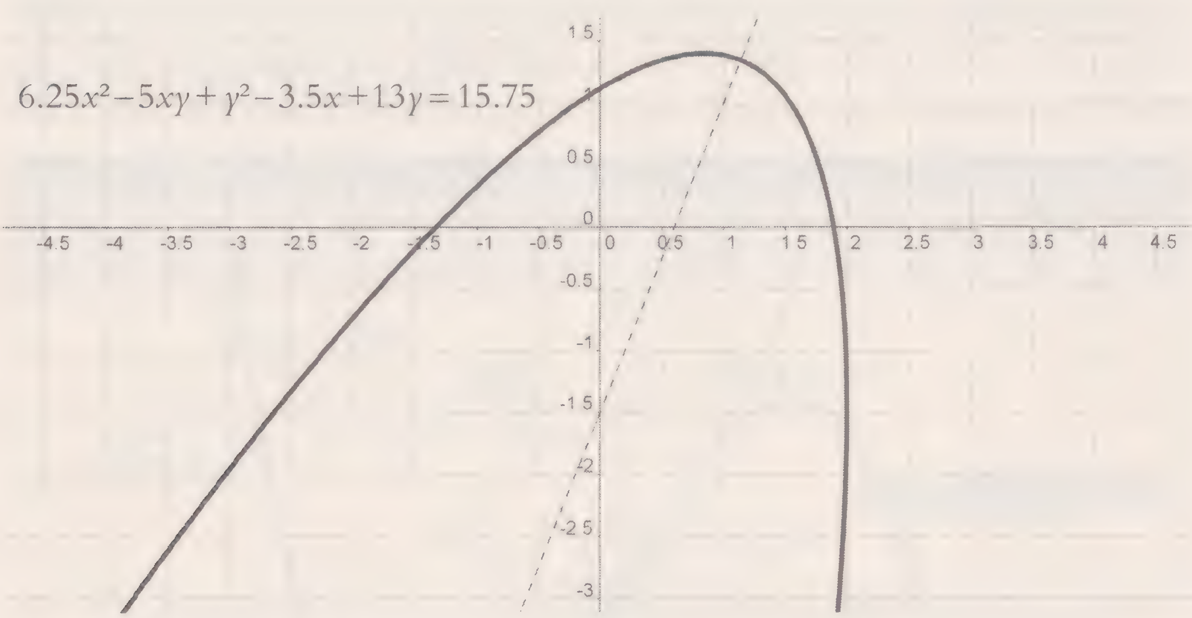
Domain of a function

The domain of a function is the set of values of the variable x in which this function exists. It is said that a function does not exist in the real field when it acquires

'strange' values, such as infinity ($+\infty$ or $-\infty$), or when the value of y is a number like $\sqrt{-16}$, which no longer falls within the field of numbers used so far (which in mathematics are the so-called real numbers). Effectively, no known real number exists that equals $\sqrt{-16}$. Why? Because if this number did exist, on raising it to its squared value (in other words multiplying it by itself) it would result in -16. But multiplying a number by itself (whether it is positive or negative) always gives a positive result, and never a negative number such as -16:

$$(-4)^2 = (-4)\cdot(-4) = +16 \text{ and } (+4)^2 = (+4)\cdot(+4) = +16.$$

Polynomial and exponential functions have their domain throughout the field of real numbers. It is written mathematically as $D(\text{polynomial functions}) = \mathbb{R}$. The figures demonstrate a second-degree polynomial function:

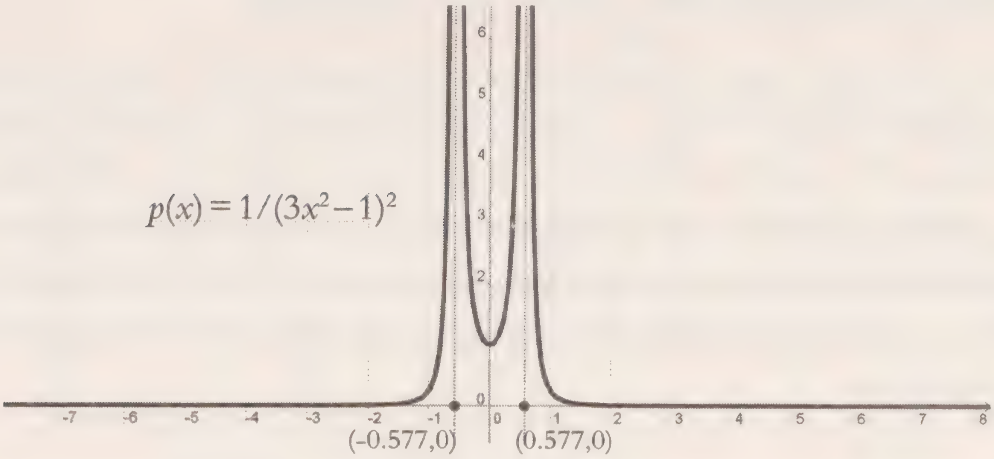


Inclined second degree polynomial curve.

Although it may appear that the curves of polynomial functions have been seen to have a beginning and an end, in reality these curves continue their shape indefinitely. Rational functions such as

$$y = \frac{1}{(3x^2 - 1)^2}$$

have some points which the function 'soars'. In other words, they tend to infinity, as can be seen in the following curve:

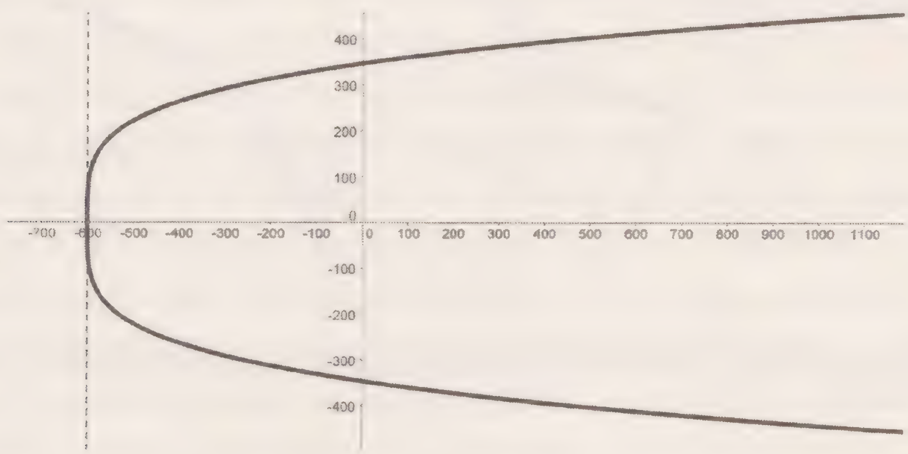


In mathematics it is often said that the function does not exist at such points and, therefore, the domain of this type of function is the entire field of real numbers minus those values of x where the function 'soars'. This is often written mathematically as:

$$D\left(\frac{1}{(3x^2-1)^2}\right) = \mathbb{R} - \left\{x = +\frac{1}{\sqrt{3}},\ x = -\frac{1}{\sqrt{3}}\right\}.$$

The logarithmic function does not exist for negative values of x as can be observed in the curve on page 55. It tends to be written: $D(\ln(x)) = \mathbb{R}^+$, which means that its domain is only positive real numbers.

The domain of the irrational function $y = 70\sqrt[4]{x+600}$ comprises all real numbers except the values of x lower than -600 (in other words, $-601, -602, -603$, etc...) as can be seen in the following curve:

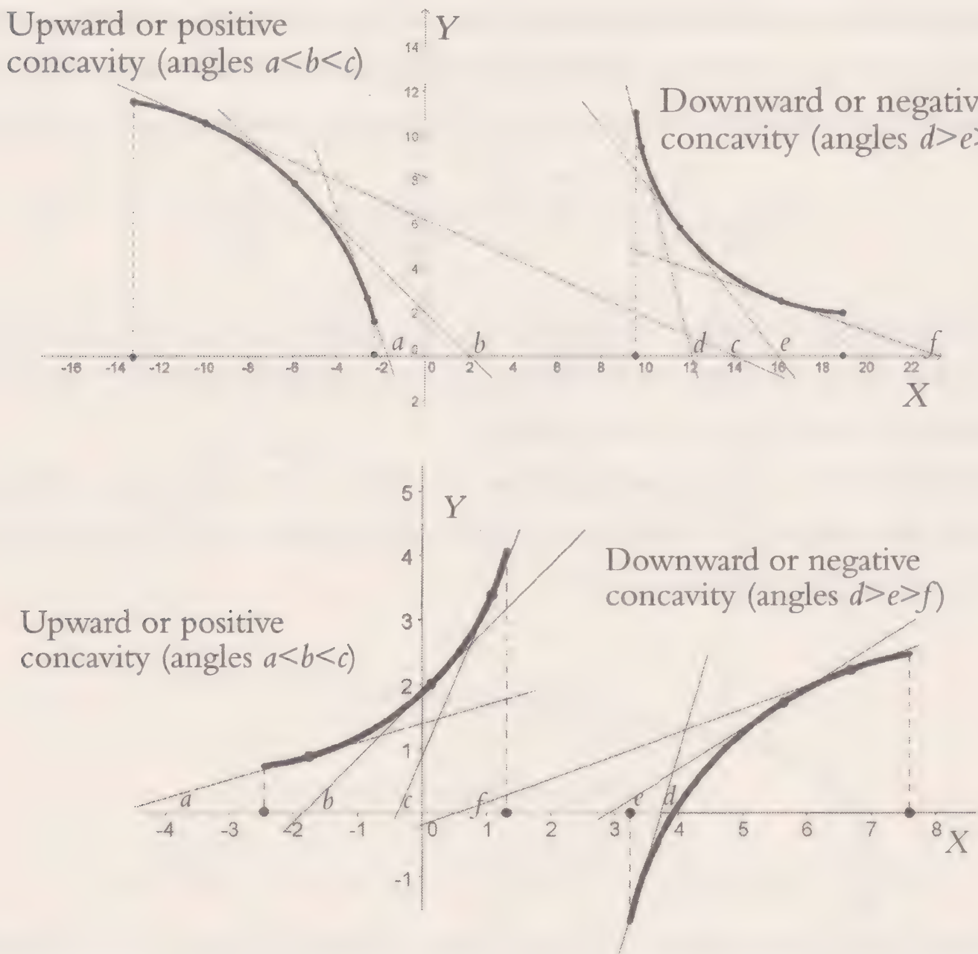


And its domain is expressed mathematically in the following way:

$$D(70\sqrt[4]{x+600}) = \mathbb{R} - \{x < -600\}.$$

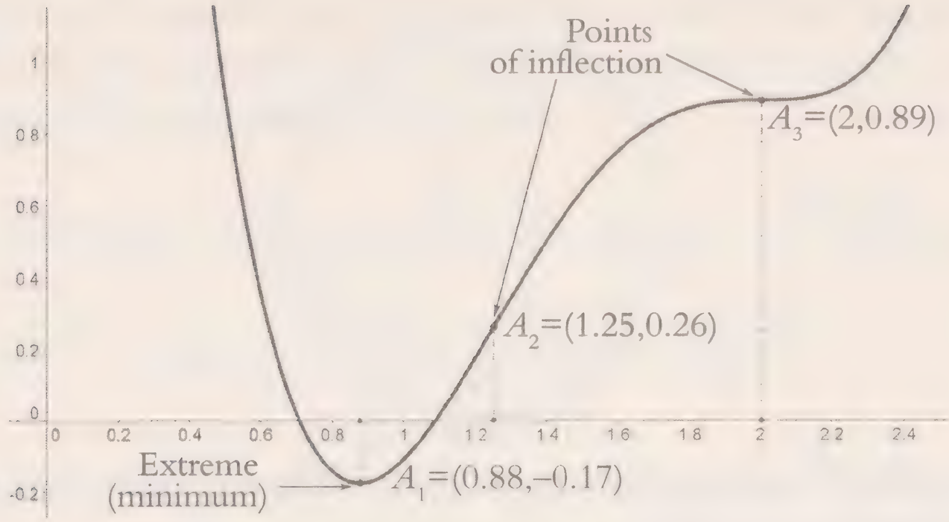
Concavity and inflection points of a function

In the sections of a curve in which there is an increase in the values themselves, two things can occur: the rate of growth keeps increasing or keeps decreasing. In the sections of a curve in which there is a decrease in the values themselves, two things can also occur: the rate of decrease can go up or down. In both cases if the rate increases this forms an upwards or positive section of the concave curve; the opposite is a downward or negative section of concavity. This can be seen in the following figures:



Two sections decreasing and two sections increasing of a curve with different concavity.

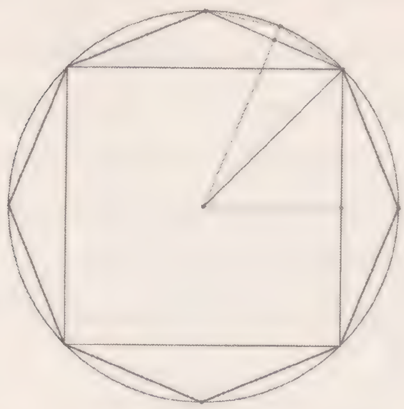
When a curve changes from a section with positive concavity to a negative concavity, or the reverse, the points with a change of tendency are called points of inflection.



Other characteristics of functions appear when drawing the curves that represent them in a graphic form, such as their circuit, curvature, normal lines (perpendicular to the curve at a point), etc.

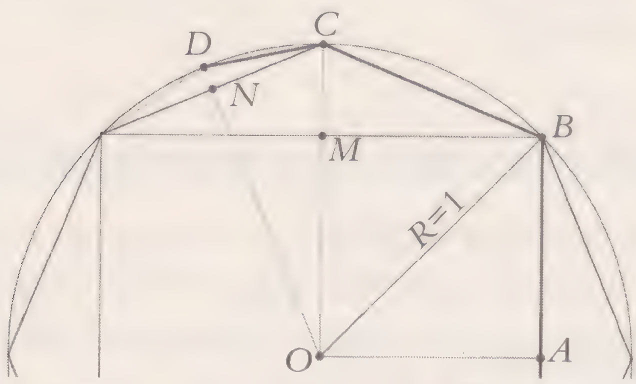
How is the length of a section of a curve measured?

The first attempts to determine the length of a curve were made by Archimedes of Syracuse in the 3rd century BC, and he began with the circle. The method used consisted of drawing two enormous regular polygons of 96 sides, one inscribed and the other circumscribed to a circle. By measuring the perimeters of both polygons he assumed that the length of the circle drawn would be somewhere between both values. This method, known as 'method of depletion', was the basis that Fermat, Newton and Leibniz used in the 17th century to develop infinitesimal calculus. This great idea of Archimedes' resulted in a brilliant advance from what was subsequently the great discovery of the scholars of the 17th century, the 'approaching the limit' method upon which all mathematics known currently as differential and integral calculus was based. In the following figures and calculations we attempt to reproduce the process carried out by Archimedes, using the current conventions of mathematical notation. The calculation that we are going to do involves obtaining the length of a circumference with a unit radius (1 cm, 1 m, 1 km...). Nowadays, we know that the result is approximately 2π, and that this approximation will be greater the more sides we have in the polygon. In the following diagram, we can see the circle, the inscribed square (regular polygon of 4 sides), the inscribed octagon (regular polygon of 8 sides) and the inscribed hexadecane (regular polygon of 16 sides).



On enlarging it we can observe the polygons of 4, 8 and 16 sides. The measurements of their sides are called:

$$AB = BM = \frac{l_4}{2}, \text{BC} = l_8 \text{ and CD} = l_{16}.$$



Right-angled triangles OAB, BMC and CND are considered, in which

$$AB = BM = OM = \frac{l_4}{2};\ BC = l_8;\ CN = \frac{l_8}{2};\ CD = l_{16}.$$

Taking radius $OB = 1$ and applying Pythagoras' theory to the triangle OAB we will have:

$$1^2 = \left(\frac{l_4}{2}\right)^2 + \left(\frac{l_4}{2}\right)^2;\ 1^2 = 2\left(\frac{l_4}{2}\right)^2;\ 2\cdot\frac{l_4^2}{4} = 1;\ \frac{l_4^2}{2} = 1;\ l_4^2 = 2;\ l_4 = \sqrt{2}.$$

Applying Pythagoras' theory once again to the right-angled triangle BMC gives:

$$l_8^2 = \left(\frac{l_4}{2}\right)^2 + \left(1 - \frac{l_4}{2}\right)^2;\ l_8^2 = \frac{l_4^2}{4} + 1 - l_4 + \frac{l_4^2}{4};\ l_8^2 = \frac{2}{4} + 1 - \sqrt{2} + \frac{2}{4} = 2 - \sqrt{2};\ l_8 = \sqrt{2 - \sqrt{2}}.$$

Applying Pythagoras' theory once again to the right-angled triangle *CND*, we must first calculate *ND*, which is $1-ON$. Therefore, *ON* must be calculated beforehand in the right-angled triangle *ONC*.

$$ON^2 = OC^2 - CN^2;\ \ ON^2 = 1^2 - \left(\frac{l_8}{2}\right)^2;\ \ ON^2 = 1^2 - \left(\frac{\sqrt{2-\sqrt{2}}}{2}\right)^2;$$

$$ON = \sqrt{1-\frac{2-\sqrt{2}}{4}} = \sqrt{\frac{4-2+\sqrt{2}}{4}} = \frac{\sqrt{2+\sqrt{2}}}{2};$$

$$ND = 1-\frac{\sqrt{2+\sqrt{2}}}{2} = \frac{2-\sqrt{2+\sqrt{2}}}{2};\ \ l_{16}^2 = CN^2 + ND^2 = \left(\frac{\sqrt{2-\sqrt{2}}}{2}\right)^2 +$$

$$+\left(\frac{2-\sqrt{2+\sqrt{2}}}{2}\right)^2 = 2-\sqrt{2+\sqrt{2}};\ \ l_{16} = \sqrt{2-\sqrt{2+\sqrt{2}}}.$$

And it can be deduced that:

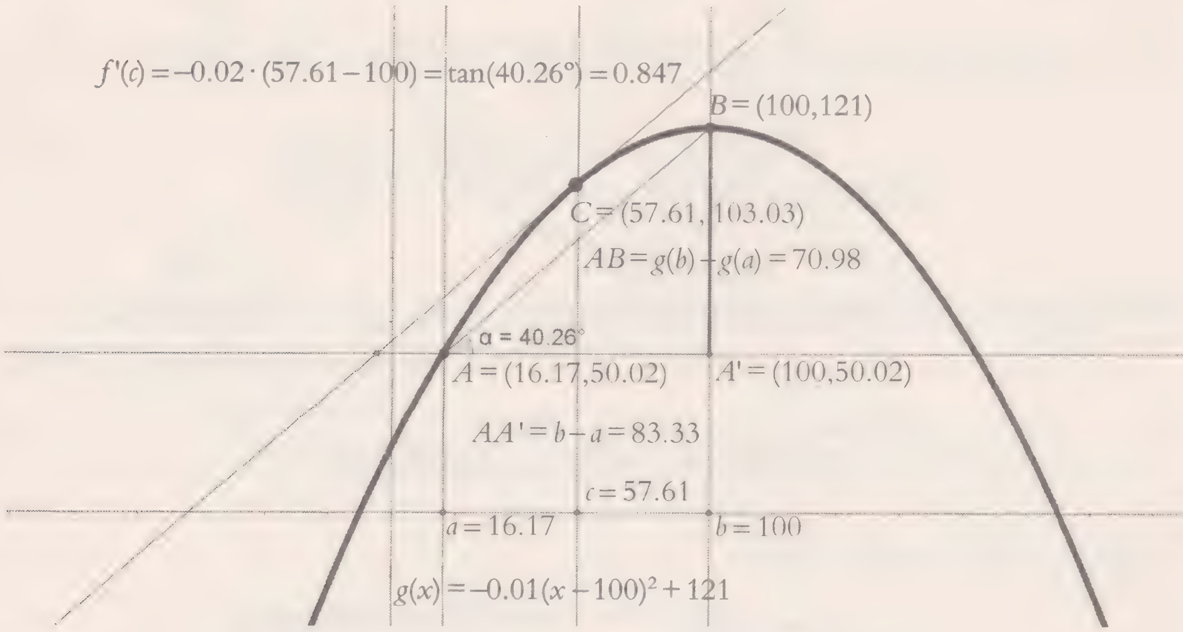
$$l_{32} = \sqrt{2-\sqrt{2+\sqrt{2+\sqrt{2}}}};\ \ l_{64} = \sqrt{2-\sqrt{2+\sqrt{2+\sqrt{2+\sqrt{2}}}}};\ldots$$

The perimeter of each polygon is calculated by multiplying the length of a side by the number of sides. Half of the perimeter approaches the value of π.

No. of sides	Side	Perimeter	$\frac{\text{Perimeter}}{2}$ = = *approximate value of* π
4	$\sqrt{2}$	$4\sqrt{2}$	2.828427125
8	$\sqrt{2-\sqrt{2}}$	$8\sqrt{2-\sqrt{2}}$	3.061467459
16	$\sqrt{2-\sqrt{2+\sqrt{2}}}$	$16\sqrt{2-\sqrt{2+\sqrt{2}}}$	3.121445152
32	$\sqrt{2-\sqrt{2+\sqrt{2+\sqrt{2}}}}$	$32\sqrt{2-\sqrt{2+\sqrt{2+\sqrt{2}}}}$	3.136548491
64	$\sqrt{2-\sqrt{2+\sqrt{2+\sqrt{2+\sqrt{2}}}}}$	$64\sqrt{2-\sqrt{2+\sqrt{2+\sqrt{2+\sqrt{2}}}}}$	3.140331157

Later on, the German mathematician Bernhard Riemann (1826-1866) applied a new concept to calculate the length of a section of the curve. He began his average

value theorem by saying that in a stretch of a continuous curve between two points A and B there exists an intermediate point C, the tangent of which is parallel to the line AB:

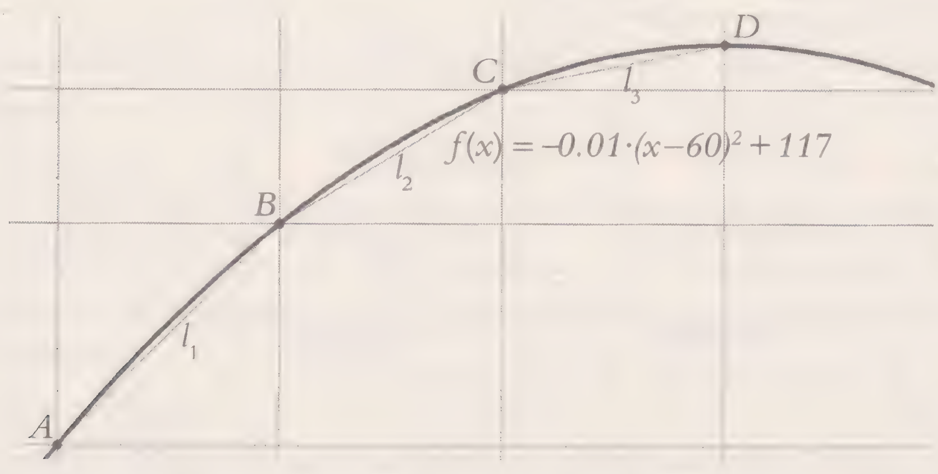


In other words, as the tangent at C is the derivative at point $C, f'(c)$ a point C should exist between A and B such that:

$$f'(c) = \frac{f(b) - f(a)}{b - a}.$$

In the specific case of this figure it can be seen that:

$$f'(57.61) = -0.02 \cdot (57.61 - 100) = 0.847 = \frac{f(100) - f(16.17)}{100 - 16.17} = \frac{70.98}{83.83} = 0.847.$$



Length of a section of a curve divided into three segments.

To calculate the length of the arc of the curve $L(ABCD)$, an initial approximation is made through three straight segments whose total length is $l_1 + l_2 + l_3$. Applying the theorem of Pythagoras to each section, we have:

$$l_1 + l_2 + l_3 =$$
$$= \sqrt{[f(b)-f(a)]^2+(b-a)^2} + \sqrt{[f(c)-f(b)]^2+(c-b)^2} + \sqrt{[f(d)-f(c)]^2+(d-c)^2}.$$

If we apply the theorem of the previously explained average value to each section AB, BC and CD:

$$[f(b)-f(a)] = f'(m)\cdot(b-a)$$
$$[f(c)-f(b)] = f'(n)\cdot(c-b)$$
$$[f(d)-f(c)] = f'(p)\cdot(d-c)$$

where m, n and p are intermediate abscissas of the segments AB, BC and CD.

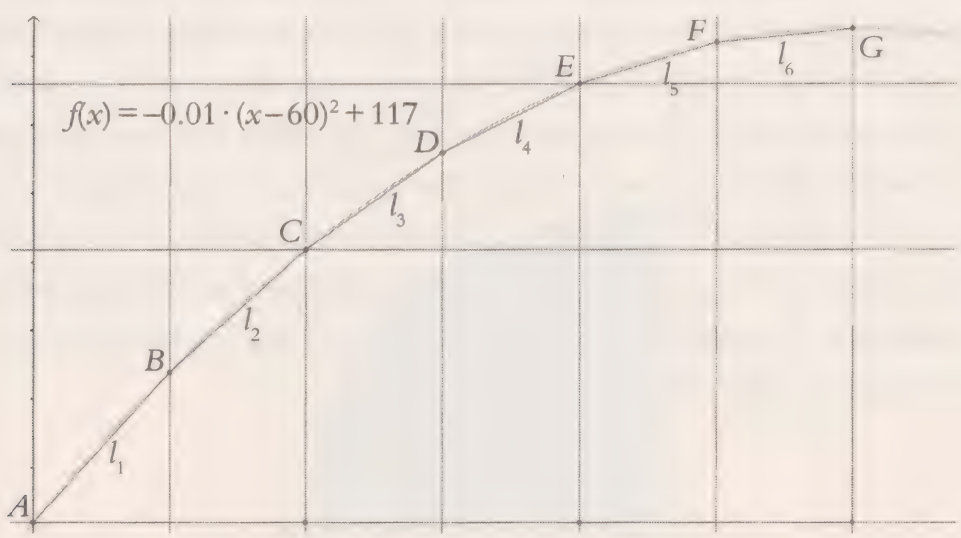
If sections of equal width are taken, $(b-a)=(c-b)=(d-c)=\Delta x$:

$$L \cong l_1+l_2+l_3 = \sqrt{f'(m)^2\cdot\Delta x^2+\Delta x^2} + \sqrt{f'(n)^2\cdot\Delta x^2+\Delta x^2} + \sqrt{f'(p)\cdot\Delta x^2+\Delta x^2}.$$

If the number of sections is increased from 3 to n:

$$L \cong l_1 + l_2 + l_3 + \ldots + l_n = \sum_1^n \sqrt{f'(x_i)^2\cdot\Delta x^2+\Delta x^2} = \sum_1^n \sqrt{f'(x_i)^2+1}\cdot\Delta x.$$

The number of segments can be increased and then the sum of its lengths are more approximate to the length of the curve section.



Length of a section of a curve divided into six segments.

If the number of sections is increased indefinitely, the sum $l_1 + l_2 + l_3 \ldots + l_n$ increasingly approximates the length L of the arc of the curve, and the length Δx is increasingly smaller, tending to zero. And when Δx is already very close to zero it is called a differential of x *(dx)*.

The limit of these sums when Δx is decreasing and there are increasingly more summands (n tends to ∞) is known in mathematics as the defined integral of the function $\sqrt{f'(x)^2+1}$ between the points of abscissa a and d:

$$L = \lim_{n\to\infty} (l_1 + l_2 + l_3 + \ldots + l_n) = \lim_{n\to\infty} \sum_1^n \sqrt{f'(x_i)^2+1} \cdot \Delta x = \int_a^d \sqrt{f'(x)^2+1} \cdot dx.$$

Rectifiable curves exist with exact calculable lengths. But there are non-rectifiable curves. Rectifiable curves are circles, parabolae, catenaries, cycloids, logarithmic spirals, straight lines… In contrast, an ellipse is non-rectifiable, because when calculating its length with the above integral formula, an integral called the elliptic integral of the second degree is obtained, which can only be approximated rather than calculated exactly. Some notable approximations have been proposed for the length of the ellipse, such as by the Indian mathematician Srinivasa Ramanujan (1887-1920):

$$L\,(\text{ellipse}, a,\ b) \ = \pi[3(a+b) - \sqrt{(3a+b)(a+3b)}\,].$$

In 1903 Ramanujan invented an algorithm which in 1987 permitted the calculation of the number π with 100 million decimal digits after a quest of two and a half millennia.

The calculation of more and more digits of π is an endless task for humans – and their computers. However, this infinite career does not seem very reasonable, given that 39 digits of π are enough to calculate the perimeter of a circle that encompasses the entire known universe with an error less that the radius of a hydrogen atom.

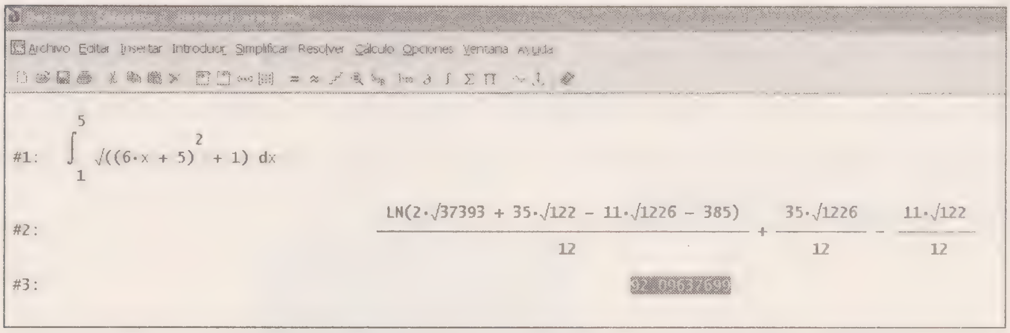
In the case of a rectifiable curve such as the parabola, the length of its arc can be calculated. For example, the length of a parabolic curve $y = 3x^2 + 5x - 10$ between the abscissa points $x = 1$ and $x = 5$ would be calculated in the following way: in this case the derivative of the function is $f'(x) = 6x + 5$, and the length between 1 and 5 can be obtained by calculating the integral

$$\int_1^5 \sqrt{(6x+5)^2+1} \cdot dx$$

by using the symbolic calculus software Derive, resulting in:

$$L=\int_1^5 \sqrt{(6x+5)^2+1}\,dx=\int_1^5 \sqrt{36x^2+60x+26}\,dx=$$

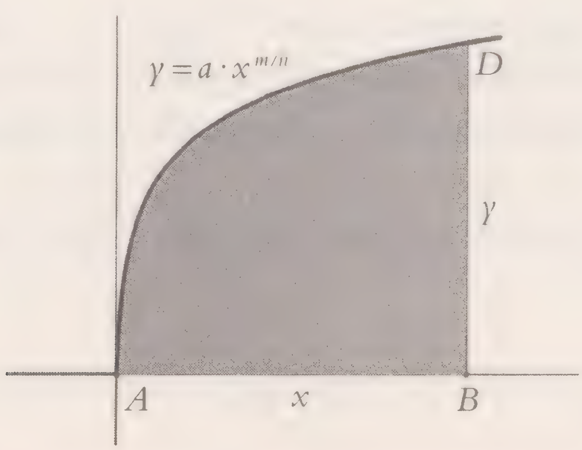
$$=\frac{\ln(2\sqrt{37393}+35\sqrt{122}-11\sqrt{1226}-385)}{12}+\frac{35\sqrt{1226}}{12}-\frac{11\sqrt{122}}{12}\cong 92.09637699.$$



How is the area enclosed by a curve determined?

In 1711 the book *De Analysi* appeared; it had been written by Isaac Newton in 1669. Fluxions appeared in it for the first time. Newton sent the book to Isaac Barrow, his mathematics professor at Trinity College, Cambridge, who gave it a very positive appraisal. The first rule to calculate quadratures (or irregular areas) found in *De Analysi* is:

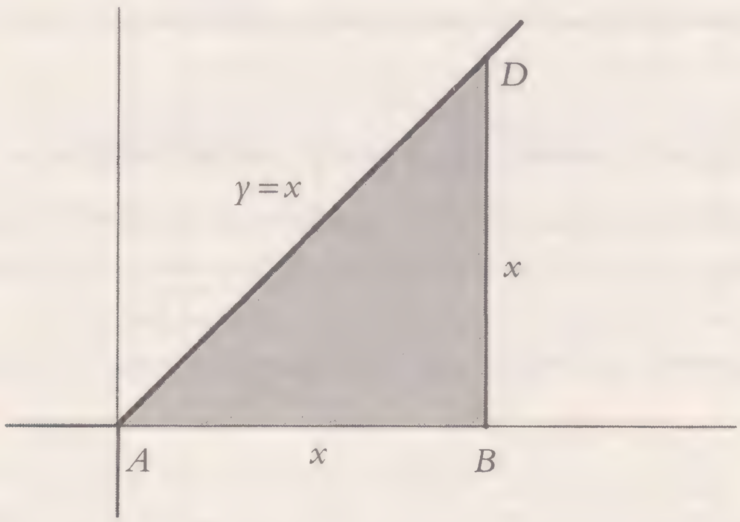
> "Let us suppose that the base AB of any curve AD has a perpendicular ordinate BD, and let us call $AB=x$, $BD=y$, and if we suppose that $a, b, c,\ldots$ are given quantities and m, n, whole numbers. Then:
>
> Rule 1: If $y=a\cdot x^{m/n}$, then $\frac{an}{m+n}\cdot x^{(m+n)/n}=$ area ABD."



Using the figure on the previous page, Newton calculated the area above the horizontal axis encompassed by the curve $y = a \cdot x^{m/n}$ up to the abscissa point x. According to Newton, and as stated in a 1745 translation of *De Analysi*, this area was:

$$\frac{an}{m+n} \cdot x^{(m+n)/n}.$$

With this statement, Newton began to develop what is currently known as integral calculus. This discovery is applied to the line $y = x$, where $m = n = a = 1$, the above formula gives us $1/2\,x^2$, which is easily confirmed with the formula of the area of the triangle: $\frac{\text{base} \times \text{height}}{2}$.



In a similar fashion, the area below a parabola $y = x^2$ between the origin and abscissa point x is:

$$\frac{x^{2+1}}{2+1} = \frac{x^3}{3}.$$

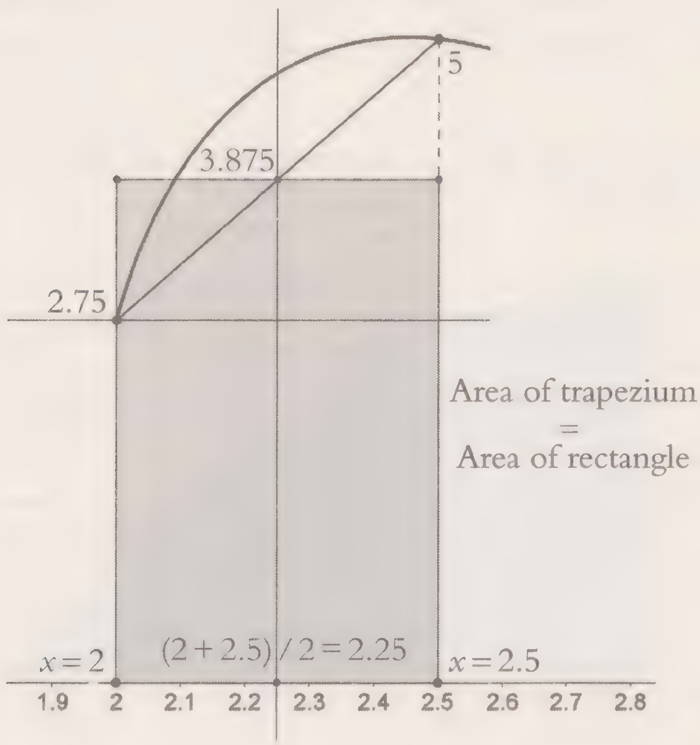
In rule 2 of *De Analysi* it says:

"If the value of y can be made equal to a number of these terms, the area may also comprise the areas that result from each one of these terms."

As an example, Newton demonstrated that the area below the curve $y = x^2 + x^{3/2}$ is:

$$\frac{x^3}{3} + \frac{2}{5} \cdot x^{5/2}.$$

Stated in current terms, the integral of a sum of functions is the sum of the integrals of each one of them. Riemann developed a systematic method to calculate the area between a curve and the axis *OX*. To do this he divided the interval between the minimum and maximum abscissas in areas in a trapezoidal form with equal length from the base to the upper points, the ordinates corresponding to the curve, as in the following figure:



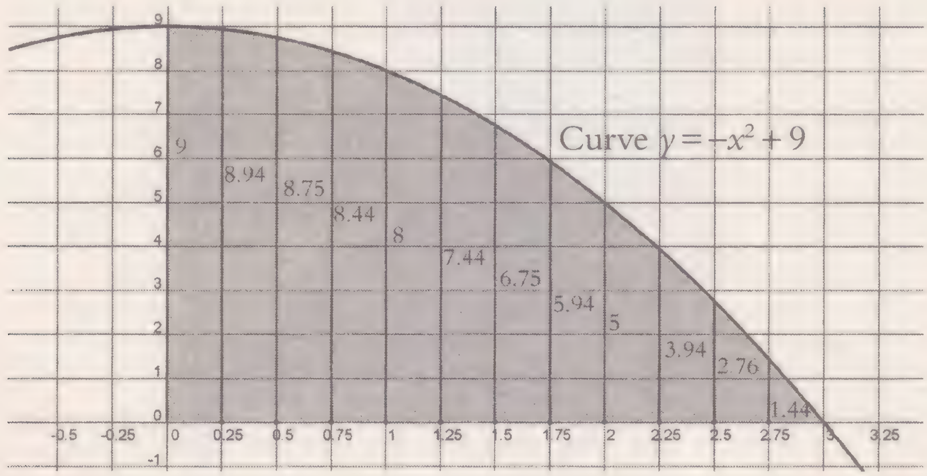
As an example we are going to calculate the area under the function $y = -x^2 + 9$ between $x = 0$ and $x = 3$, considering four approximations and employing four different divisions in trapezoidal widths 1, 0.5, 0.25 and 0.125. With divisions of width 1 an approximation of the area is obtained between $x = 0$ and $x = 3$ of 17.5 units squared.

x_1	x_2	$f(x_1)$	$f(x_2)$	$(f(x_2)+f(x_1))/2$	Δx	$(f(x_2)+f(x_1)/2)\cdot\Delta x$
2	3	5	0	2.5	1	2.5
1	2	8	5	6.5	1	6.5
0	1	9	8	8.5	1	8.5
						17.5

With divisions of width 0.5 an approximation of the area is obtained of 17.875 units squared.

x_1	x_2	$f(x_1)$	$f(x_2)$	$(f(x_2)+f(x_1))/2$	Δx	$(f(x_2)+f(x_1)/2)\cdot\Delta x$
2.5	3	2.75	0	1.375	0.5	0.6875
2	2.5	5	2.75	3.875	0.5	1.9375
1.5	2	6.75	5	5.875	0.5	2.9375
1	1.5	8	6.75	7.375	0.5	3.6875
0.5	1	8.75	8	8.375	0.5	4.1875
0	0.5	9	8.75	8.875	0.5	4.4375
						17.875

With divisions of width 0.25 an approximation of the area is obtained of 17.984375 units squared:



x_1	x_2	$f(x_1)$	$f(x_2)$	$(f(x_2)+f(x_1))/2$	Δx	$(f(x_2)+f(x_1)/2)\cdot\Delta x$
2.75	3	1.4375	0	0.71875	0.25	0.1796875
2.5	2.75	2.75	1.4375	2.09375	0.25	0.5234375
2.25	2.5	3.9375	2.75	3.34375	0.25	0.8359375
2	2.25	5	3.9375	4.46875	0.25	1.1171875
1.75	2	5.9375	5	5.46875	0.25	1.3671875
1.5	1.75	6.75	5.9375	6.34375	0.25	1.5859375
1.25	1.5	7.4375	6.75	7.09375	0.25	1.7734375
1	1.25	8	7.4375	7.71875	0.25	1.9296875
0.75	1	8.4375	8	8.21875	0.25	2.0546875
0.5	0.75	8.75	8.4375	8.59375	0.25	2.1484375
0.25	0.5	8.9375	8.75	8.84375	0.25	2.2109375
0	0.25	9	9.0625	9.03125	0.25	2.2578125
						17.984375

With divisions of width 0.125 a better approximation of the area is obtained, 17.9921875 units squared.

x_1	x_2	$f(x_1)$	$f(x_2)$	$(f(x_2)+f(x_1))/2$	Δx	$(f(x_2)+f(x_1)/2)\cdot\Delta x$
2.875	3	0.734375	0	0.3671875	0.125	0.045898438
2.75	2.875	1.4375	0.734375	1.0859375	0.125	0.135742188
2.625	2.75	2.109375	1.4375	1.7734375	0.125	0.221679688
2.5	2.625	2.75	2.109375	2.4296875	0.125	0.303710938
2.375	2.5	3.359375	2.75	3.0546875	0.125	0.381835938
2.25	2.375	3.9375	3.359375	3.6484375	0.125	0.456054688
2.125	2.25	4.484375	3.9375	4.2109375	0.125	0.526367188
2	2.125	5	4.484375	4.7421875	0.125	0.592773438
1.875	2	5.484375	5	5.2421875	0.125	0.655273438
1.75	1.875	5.9375	5.484375	5.7109375	0.125	0.713867188
1.625	1.75	6.359375	5.9375	6.1484375	0.125	0.768554688
1.5	1.625	6.75	6.359375	6.5546875	0.125	0.819335938
1.375	1.5	7.109375	6.75	6.9296875	0.125	0.866210938
1.25	1.375	7.4375	7.109375	7.2734375	0.125	0.909179688
1.125	1.25	7.734375	7.4375	7.5859375	0.125	0.948242188
1	1.125	8	7.734375	7.8671875	0.125	0.983398438
0.875	1	8.234375	8	8.1171875	0.125	1.014648438
0.75	0.875	8.4375	8.234375	8.3359375	0.125	1.041992188
0.625	0.75	8.609375	8.4375	8.5234375	0.125	1.065429688
0.5	0.625	8.75	8.609375	8.6796875	0.125	1.084960938
0.375	0.5	8.859375	8.75	8.8046875	0.125	1.100585938
0.25	0.375	8.9375	8.859375	8.8984375	0.125	1.112304688
0.125	0.25	8.984375	8.9375	8.9609375	0.125	1.120117188
0	0.125	9	8.984375	8.9921875	0.125	1.124023438
						17.9921875

Calculating the area using the definite integral from Newton's quadrature method described above, we get:

$$\text{Area} = \int_0^3 (-x^2+9)\,dx = \left[\frac{-x^3}{3}+9x\right]_0^3 = \left(\frac{-3^3}{3}+9\cdot 3\right)-\left(\frac{-0^3}{3}+9\cdot 0\right) = -9+27 = 18.$$

And an increasing approximation can be seen in each of the trapezoidal divisions up to the exact value of the area, which is 18.

Chapter 3

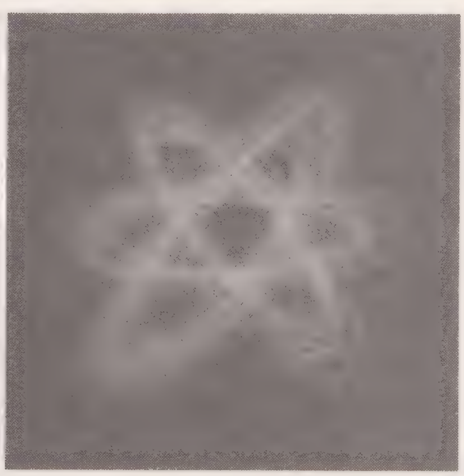
The Paths of Curves: Trajectories of Objects

Trajectories, movement and curves

The trajectory is the path that an object follows in its movement. It can be defined as the set of all the positions that a given object occupies in its movement. It is possible to define trajectories from elemental particles (electrons, photons...) up to objects of a greater size (missiles, cars, satellites, planets...) and even fluids (water, stellar material, hurricanes...); the common element of all trajectories is their movement.

The form of trajectories can be straight, curved, cyclic...There are also continuous and non-continuous trajectories. The ones we can see are continuous, associated with the movement of an object, but if trajectories of atomic objects are defined, when quantum mechanics comes into play, these can be non-continuous. Quantum mechanics is the part of physics that studies the movement of objects of the size of atomic nuclei. An example of such a trajectory is one that follows an electron surrounding the atomic nucleus; in this case we talk of probabilistic trajectories. These are 'zones' of different probabilities of movement and not exactly a definite line of a trajectory.

Trajectory of a space rocket.

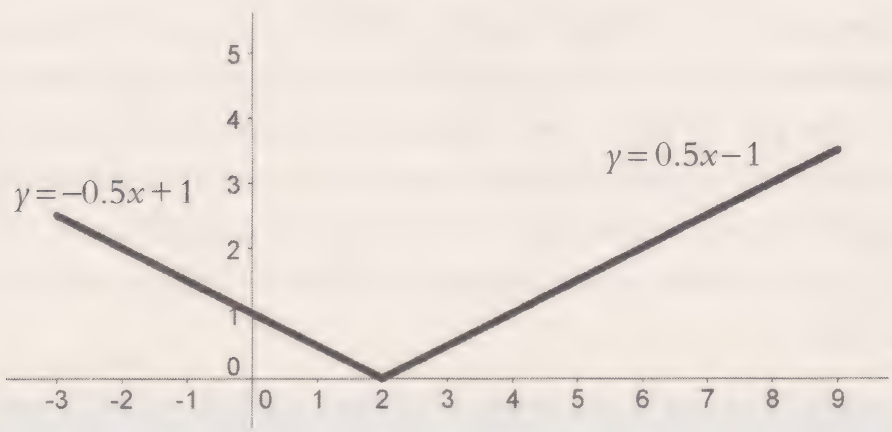


Trajectory of an electron.

A curve converts into a line

We are going to study the trajectory followed by an object that falls on Earth and that of a bullet over a short distance. In the case of the object, the graph representation is a vertical line. In the case of a bullet, the trajectory that it follows is a straight line if short distances are considered, while over long distances the trajectory of a bullet is not straight, but rather it tends to curve, as the force of gravity surpasses the impulse given to the bullet by the weapon. Usually the police carry out investigations of ballistic trajectories considering them to be straight lines.

In the scenario of a shot it is possible to find out the impact point of the projectile and the ricochet off the ground taking simple measurements, and as such it is possible to calculate the trajectory of the shot and to determine the exact position of the shooter and the angle of the shot. In the following figure a simulation of a shot with a ricochet has been created to enable us to make the calculations. In their simulations the police use laser light beams, taut coloured thread or extendible tubes; all of them draw a straight line across the space.



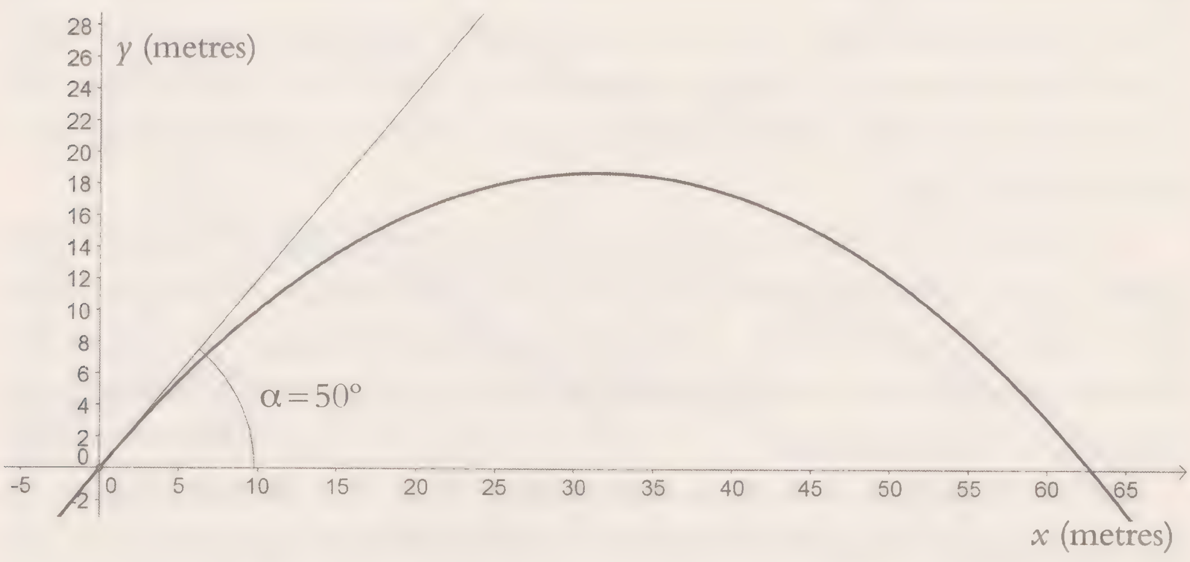
The line converts into a curve

In general, trajectories of movements tend to be definite curves or a combination of two or more curves. In the case of launching a projectile or any other object over a long distance, the trajectory is not a straight line, as stated in the previous section. The reason that this trajectory curves is due to the influence that gravity exerts over the object, air resistance and the Coriolis force (the effect of the Earth's rotation which tends to reroute the projectile towards the right of its trajectory when the

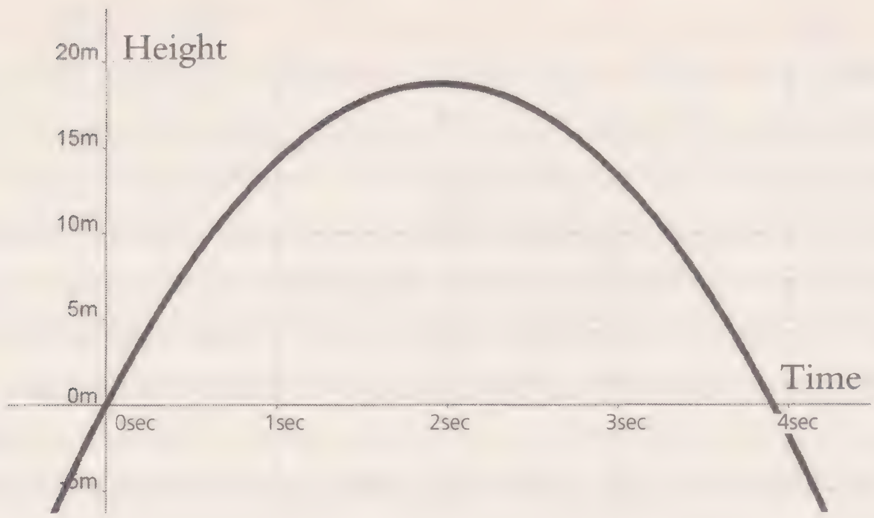
movement takes place in the northern hemisphere). If gravity alone was the only influence, the curve that would be obtained would be a parabola, but in reality the trajectory that the object follows is a curve that approximates a parabola.

Galileo was the first person who attempted to calculate a vertical fall, by dropping something from the top of the Tower of Pisa. However, this movement was too fast to make precise measurements. To 'reduce' the effect of gravity, he carried out a mythical experiment in which he dropped a ball along a plane inclined at an angle of α, which reduced the acceleration due to gravity. He repeated the test using different inclines. Using a water clock, he observed the time that it took for the ball to fall, and from analysing the data gathered he deduced the formula (equation) for movement, in other words, the height the ball was at the moment of its fall according to the time of its descent. From this data he obtained the equation that related the height with the time taken to descend. From the data obtained for different angles he deduced that the movement generated some speeds in the ball that were proportional to the time taken to descend, from which he deduced that the increase in the speed of the fall is uniform. With the data obtained for different angles he was able to calculate a value of this acceleration due to gravity over the Earth which is very close to the accepted value today: $g=9.8$ m/s^2.

After that experiment, Galileo studied the movement of projectiles in depth and contributed great advances to the artillery of his time. From the works by the Grecian Apollonius (third century BC) he deduced that the trajectory of a canon ball was a curve called a parabola which was already known in classical Greece.



Curve of the trajectory of a ball dropped at an incline of 50° and an output speed of 90 km/h. $y=1.19x-0.019x^2$.



Parabola of Galileo's throw. The height of the projectile depends on the time from the shot; it has a maximum halfway through its journey time. The curve demonstrates the height *that the ball reaches depending on the* time *that has passed since the shot, with an incline of 50° and an exit speed of 90 km/h. Height = 19.15 · time – 4.9 · time²*.

The two previous curves, although they appear to be similar, refer to very different questions and have very different uses. The curve of $y = 1.19x - 0.019x^2$ is a curve which can be seen in reality. It is the trajectory, or rather, the different positions x-y that the bullet follows through the air after the shot is fired until it reaches the ground. In contrast, the *head curve* = 19.15 · time – 4.9 · time² or, simplifying the names of the variables, $y = 19.15 \cdot t - 4.9 \cdot t^2$, is not a real curve, but rather it describes a phenomenon of science and cannot be physically seen in the firing range. If you follow the curve, each point (described by a pair of values y-t) indicates the height which the bullet reaches at every instant t of its movement, from the moment the shot is fired.

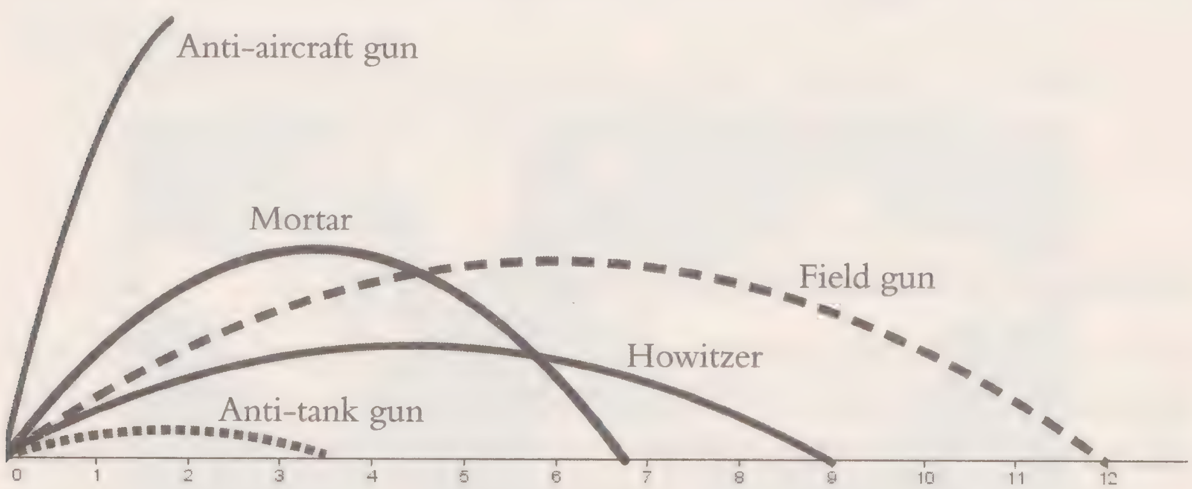
The tables on the following page, of values calculated with the formulae, also explain the two previous curves. In the trajectory table we can see that the bullet arrives on the ground (height $y = 0$) at 62.78 m from the point it is fired. In the curve height-time it can be verified that the bullet collides with the ground ($y = 0$) between 3.9 and 4.0 seconds.

Galileo elaborated some tables with parabolic shots that were used widely in wars in European up until the 17th century. When friction is taken into account with the launch, the trajectory is almost parabolic. The effect of friction is intensified with the increase in speed of the launch.

Trajectory	
x metres	*y* metres
0	0
5	5.48075
10	10.013
15	13.59675
20	16.232
25	17.91875
30	18.657
35	18.44675
40	17.288
45	15.18075
50	12.125
55	8.12075
60	3.168
62.7833	0

Curve height-time	
t seconds	*y* metres
0	0
0.3	5.3040
0.6	9.7260
0.9	13.2660
1.2	15.9240
1.5	17.7000
1.8	18.5940
2.1	18.6060
2.4	17.7360
2.7	15.9840
3	13.3500
3.3	9.8340
3.6	5.4360
3.9	0.1560
4.0	-1.8000

The launch of objects is a discipline that has been studied since ancient times for the purpose of applying it to the launch of projectiles in battles. As shown in the figure, all the trajectories form a family of parabolas dependent on two parameters: angle of the shot and speed of the exit.



Different trajectories to launch projectiles.

Objects exist whose launch does not respond to the parabolic scheme of projectiles. One example is a boomerang, an object for hunting with very old origins

and which has been used throughout the world, not only in Australia as commonly thought, although its name is a transcription of the pronunciation by Australian aborigines. Over time, mamny types of throwing weapons were developed and now they are used in sporting activities.

The classic boomerang is made up of two wings of the same length, joined together by an elbow. The wings, one to attack and the other to exit, are straight or tapered, and the angle between them may be between 105° and 110°. The blades have a similar profile to the wings of a plane, which means that the speed of the wind is greater over the upper surface than the lower surface, and therefore the pressure in the latter is greater, pushing the wing upwards. This principle is also applicable to the boomerang, making it 'fly'; in physics this is known as the Bernoulli principle.

When the boomerang is thrown it manages to rotate around its own axis and, therefore, a gyroscopic effect is created (it displaces itself and turns around) which makes the trajectory such that it returns to the point it was thrown from. The trajectory being described is more or less elliptic.

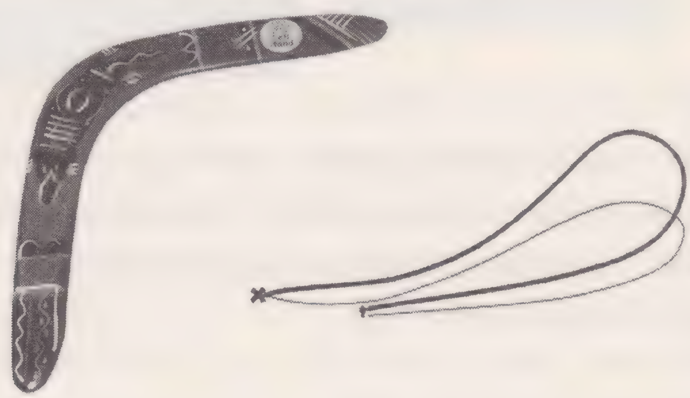
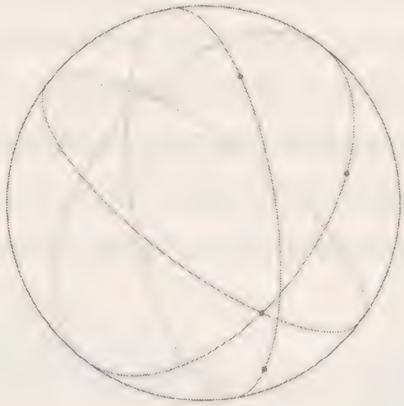


Diagram of the trajectory that a boomerang like the one in the photograph follows. The thin line is the projection of the movement on the ground.

Curves over short distances

The calculation or estimation of the shortest distance between two points is of great interest from a mathematical and physical perspective, and that of everyday life, and it encompasses, for example, the calculation of the trajectory that a vehicle equipped with GPS follows, the trajectories of international flights, or the layout of communication cables, among many other applications. The mathematical concept associated with the idea of the shortest distance between two points is the geodesic line. Although it may sound strange, a straight line is not always the shortest

distance between two points. If in the space in which one works the shortest distance between two points is a straight line, it is called a Euclidean space with a defined metric distance. The line, the plane and 3D space are Euclidean spaces of one, two and three dimensions. However, when it involves moving from one point to another over a spherical surface, the shortest line is a circumference and not a straight line.



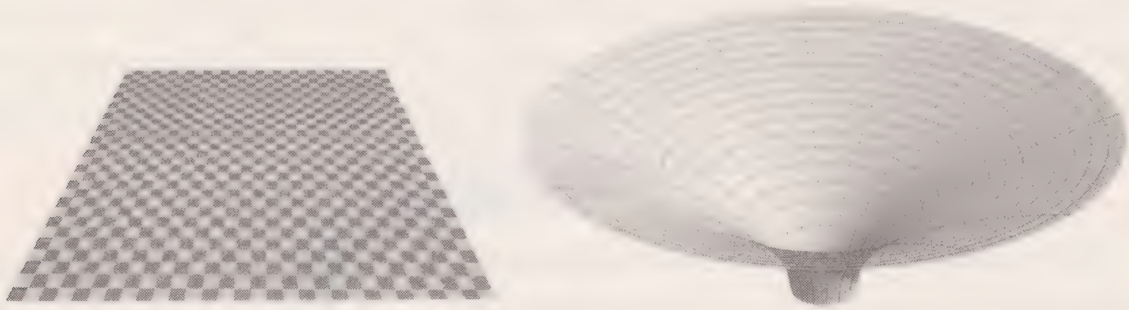
Geodesics in a sphere.

Many situations exist in which the need to calculate short routes over a spherical surface arises, as occurs with aerial routes or journeys by land over great distances (for short distances we can consider the Earth to be flat). In this case, the shortest line between two points is the arch of a circumference or a great circle. In the case of our planet, if lines are drawn corresponding to the shortest distance between two points, the geodesics, and then we draw a flat projection on paper, an image like the one that follows is obtained:

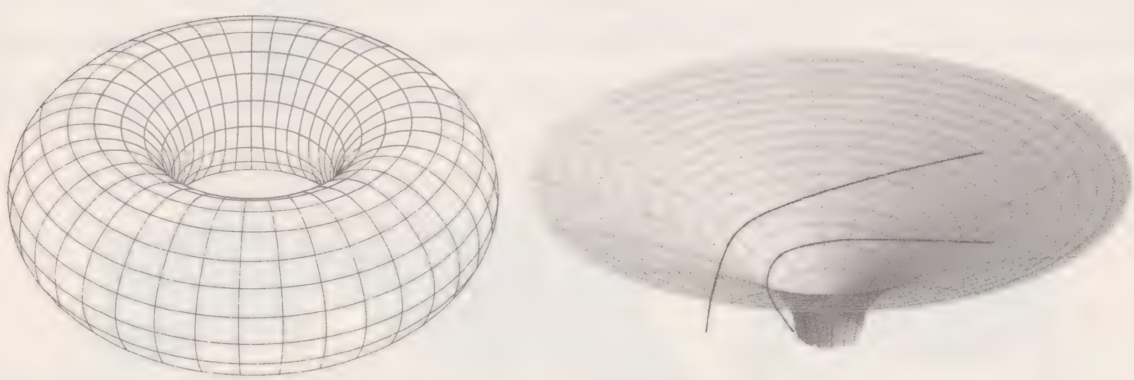
On a map of the Earth's surface, the geodesics adopt the form that can be seen on the planisphere. For example, this means that a plane that wants to travel between two points covering the shortest distance possible should do so following one of these geodesic lines.

In the case of spaces of complex curved forms, the geodesics are very important for calculating forms, minimum distances, light trajectories... They are fundamental elements in the case of the general theory of relativity and in situations of elevated gravity where strong spatial curvatures are produced.

The 'form' of a space is described mathematically with a so-called metric equation. This is the relationship between the dimensions of space (which can be represented by x, y, z) which indicate its form: it can be flat, spherical or adopt more complex forms, with different curvatures in different parts of the same form.



Flat (left) and curved space.



On the left, the curvature of a torus. The figure on the right demonstrates how the geodesic lines 'adapt' to the curved space to achieve the shortest path.

To be able to calculate the geodesics in a determined space an equation is formulated for the length of the curve in the form of the function and is calculated for the minimum value. The complexity of the process is heightened, and it is necessary to know the calculation of variations to be able to complete it.

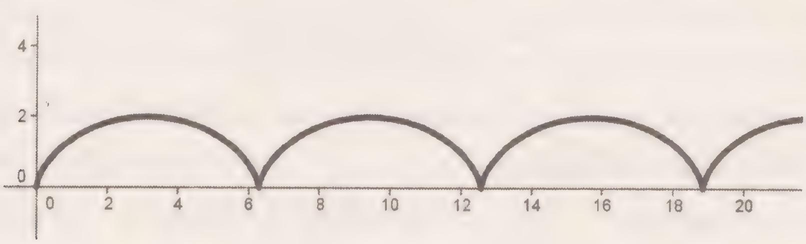
The geodesics are flow lines in the case of zones with a gravitational field, in other words, they are 'natural' routes along which the objects move under the effect of gravity. Opposite the minimum route lines we can find the lines that cover the route in the least time, known as brachistochrones.

All these geodesics are defined by the curvature of space itself, whether by the force of gravity or by the specific form of the space. The important element is the curvature of space where the object moves.

Situations can be found in which space is flat and the corresponding geodesic is not a straight line, such as in a magnetic field. In this situation, the minimum distance of the trajectory between the points of the magnetic field of a particle is a cycloid, with the condition that the magnetic and electric fields are uniform and perpendicular in themselves. We obtain a cycloid orthogonal to the magnetic field.

Moving curves: curves defined by movement

Not only are there curves that define trajectories and movement, but also others exist that are defined by the movement of geodesic figures. Take the the case of the cycloid in the figure:



The parametric equations of the cycloid have the following appearance (to make the graph the following value has been taken for $R=1$):

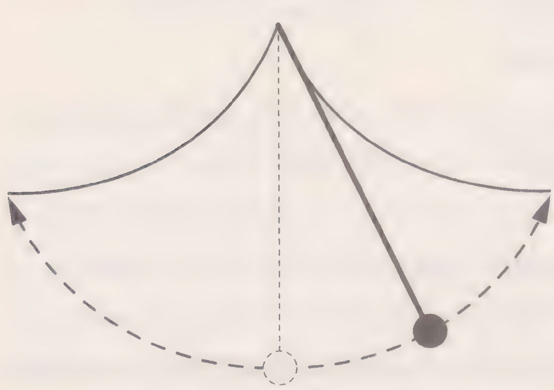
$$\begin{cases} x = R(t - \sin t) \\ y = R(1 - \cos t) \end{cases}$$

The cycloid is the curve drawn for a point on a circumference (called the generatrix circumference) when this rotates around a line (guideline) without sliding down it. Cycloids have been studied by many mathematicians due to their interesting properties. The first is that the area enclosed by a cycloid arch is exactly three times the area of the circumference that generates it; the second is that the length of the cycloid arch is four times the diameter of the circumference that generates the curve.

The great interest evoked by this curve comes from the curious characteristics it presents. Apart from the relationship between the area and the diameter, it has two curious characteristics: that of being a brachistochrone and a tautochrone. Therefore, the cycloid is the trajectory of least time in the movement of an object between two points situated on the same plane, but not on the vertical one; therefore, it is a brachistochrone. Moreover, if the cycloid is positioned looking upwards and two objects are dropped from different points, they will both arrive at the lower part at the same time. In other words, if an object is dropped that displaces itself uniformly during the fall due to the effect of gravity, it does so at the same time regardless of the initial position of the object; therefore, the cycloid is a tautochrone.

In past eras, the need and difficulty of measuring time precisely during maritime voyages forced many scientists to carry out investigations into the subject. In 1673, the discovery of the tautochronous property of the cycloid allowed Christiaan Huygens (1629-1695) to resolve the problem of constructing a clock with an isochronous pendulum. The solution consists in forcing the movement of regression of the pendulum through a buffer on each extreme point that has the same form as the cycloid curve that the end of the pendulum follows. In his work *Horologium oscillatorium sive de motu pendulorum ad horologia aptato demostrationes geometricae* (Paris, 1673) he gives a geometric demonstration of this fact. In the figure we can see that a pendulum constructed with these characteristics follows the trajectory of a cycloid whose period (time the pendulum takes to make an oscillation) results independently of the amplitude of the oscillation.

Joseph-Louis Lagrange (1736-1813) started to work on the problem of the tautochrone in 1754. In August of the following year, when he was only 19 years old, he communicated his advances in resolving the problem of the tautochrone in his correspondence to Euler, as well as his method of resolving the maximum and minimum conditions (method of multipliers). Subsequently, in 1823, Niels Henrik Abel (1802-1829) proposed a generalisation of the tautochrone problem. More specifically, he raised the issue of determining a curve so a material object that moves uniformly during a fall due to the effect of gravity and without friction through the curve up to a point given on the same, is done in a prefixed time in advance for each possible height from

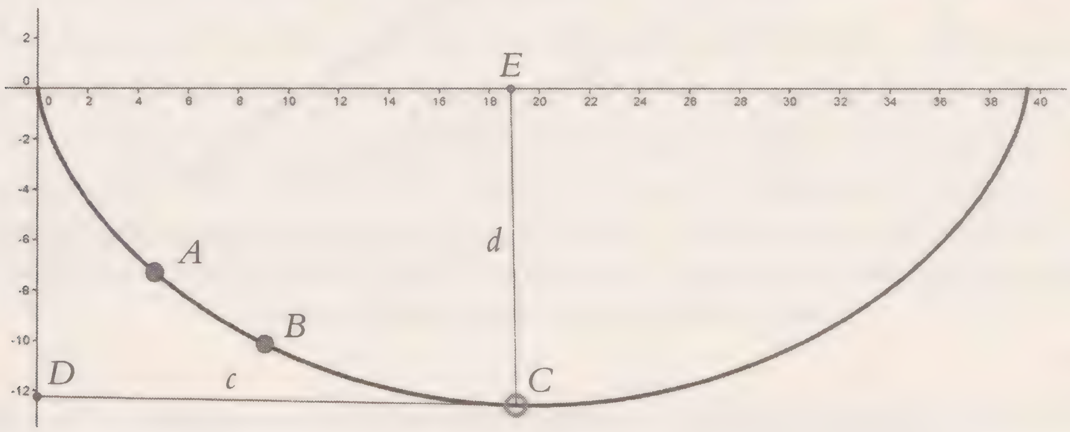


Huygens' pendulum oscillating.

which it can fall. It is evident that, if we require the time of the fall to be independent of the height from which an object has fallen, then the original problem of the tautochrone is obtained. Returning to the equations of the cycloid curve,

$$\begin{cases} x = R(t - \sin t) \\ y = R(1 - \cos t) \end{cases}$$

and supposing that A and B are two points of the cycloid in the previous figure, we can calculate the time they would take in reaching point C. To make the calculations we have to determine the distance between points A and C, and between B and C. It is supposed that the movement is produced through the cycloid and that there is no friction. The cause of the movement is solely the force of gravity.



The speed along the arch is:

$$\frac{ds}{dt} = 2R\,\sin\left(\frac{t}{2}\right).$$

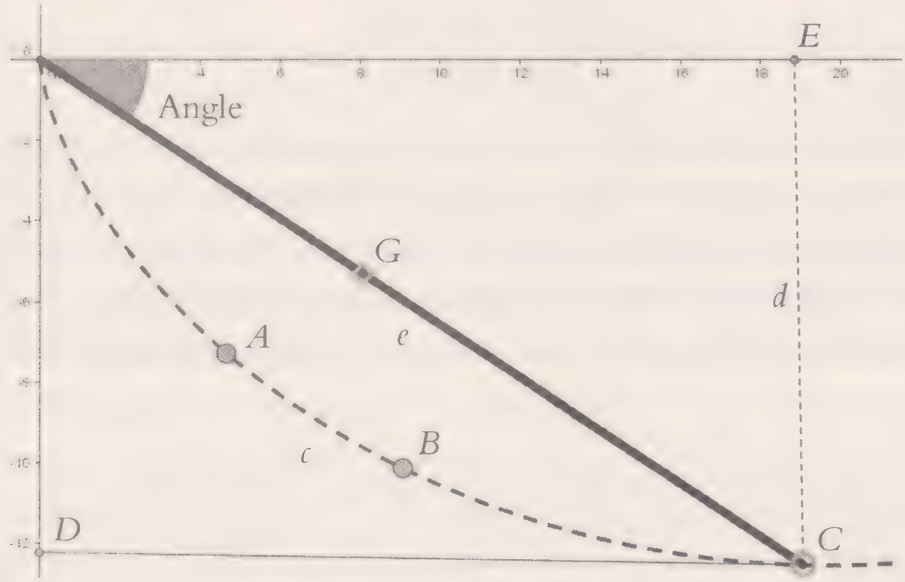
For each point of the curve and for each value of t a value for the length of the arch is obtained. In the case of the previous figure, the value of the arch from A to C is 15.36 cm, while in the case of B to C it is 10.31 cm.

The time is calculated according to the speed and position. The calculations involve some integrals, and the final result is:

$$\text{Time} = \pi\sqrt{\frac{R}{g}}.$$

The time that the object takes to arrive at point C is independent of the initial position, and solely depends on the radius R of the generatrix circumference and

the value of gravity ($g = 9.8\ \text{m/s}^2$). This result is surprising, as in the case of the cycloid, the fall time to the minimum point of the same (point C) is independent of the initial position, and it is lower than in the case of the fall over a line.



Fall of an object down a line in relation to its fall down a cycloid curve. If it descends down the line it takes longer to arrive at the bottom. Starting from any height the descending time is the same, but not the speed of the arrival at the lower part.

To make the calculations we need to know that the acceleration of an object that falls down the line is proportional to the acceleration due to gravity g multiplied by the sine of the angle indicated in the previous figure. In that situation the fall time to point C is

$$\text{Time} = \sqrt{4+\pi^2} \cdot \sqrt{\frac{R}{g}},$$

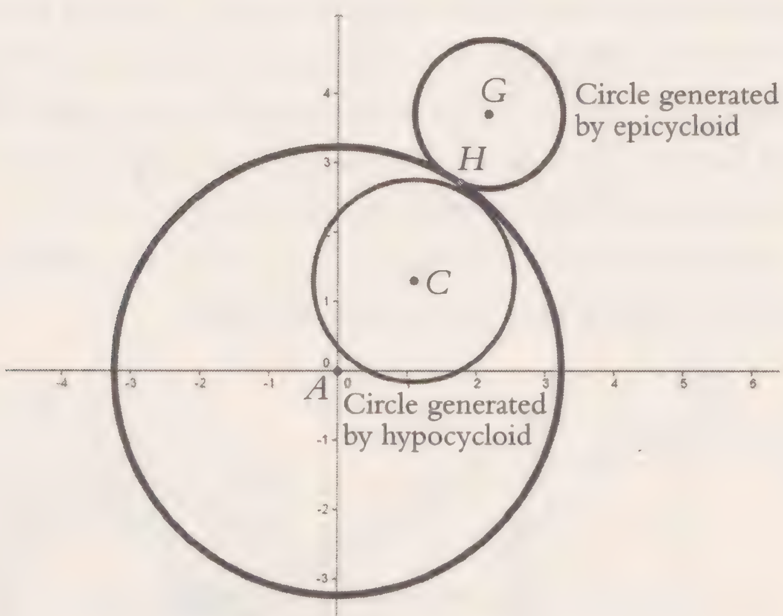
which is clearly superior to the cycloid case.

The properties of the cycloid permit the construction of a pendulum that has an oscillation period independent of the amplitude with which the movement occurs, such as Huygens' isochronous pendulum, an instrument that became essential for knowing the time on a moving ship. On ships that navigated the high seas it was possible to calculate the latitude relatively easily by using astronomic observation: the height of the Sun over the horizon at midday makes it possible to obtain the distance in degrees of latitude from the Equator; this way, measuring the height of the Sun or, in the case of the northern hemisphere, measuring the position of the Polar star during the night, it is possible to determine the latitude with certain exactitude.

On the other hand, the problem of calculating the longitude was much more complex: it was necessary to know the local time and the reference, given that the time difference between the time the Sun reached its highest point in a specific place and the time it reaches it in another can be used to obtain the angular distance between both points. But in order to do this, accurate clocks were required.

In general we can classify the curves that are constructed through movement of one curve over another figure by introducing the concept of 'roulette'. This is a curve that is formed by the trajectory from one point, called the generator, which is connected to a given curve and which spins without sliding across another fixed curve. Some examples of these types of curves are cycloids, which we have already seen, hypocycloids, epicycloids, hypotrocoids and epitrocoids.

The *hypocycloid* is the curve that describes the trajectory of a point located over a circumference that spins through the interior of another without slipping. The *epicycloid* is the curve that describes the trajectory of a point located over a circumference that spins on the exterior of another without dislodging; it is associated with planetary movement.



In the case of the hypocycloid and epicycloid, the circumference spins around another. In the case of the cycloid, the circumference spins on another circumference with an infinite radius, in other words, on a straight line. The equation of the hypocycloid has the form:

$$\begin{cases} x = (R-r)\cos\theta + r\cos\left(\dfrac{R-r}{r}\theta\right) \\ y = (R-r)\sin\theta - r\sin\left(\dfrac{R-r}{r}\theta\right) \end{cases}$$

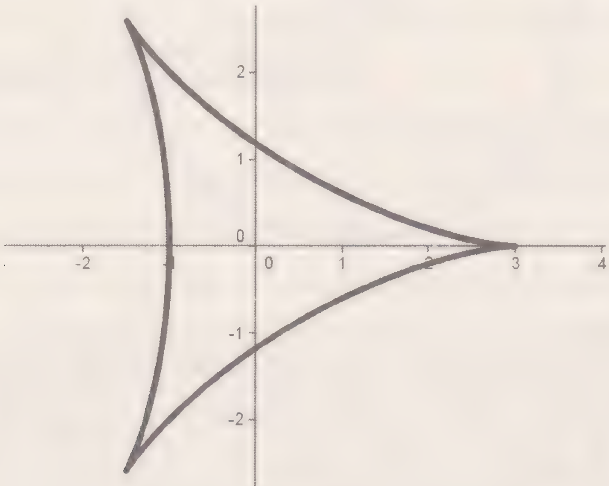
where R is the radius of the fixed circumference and r is that of the circumference that moves. The angle θ is the parameter that defines the curve and the angle that describes the movement of the circumference that moves.

An important element in curves defined by the movement of others is the relationship between their dimensions, in this case between the radii of the circumferences. This relationship will provide us with different forms of the curves. Introducing the value $K = \dfrac{R}{r}$, the equations can be written in the following way:

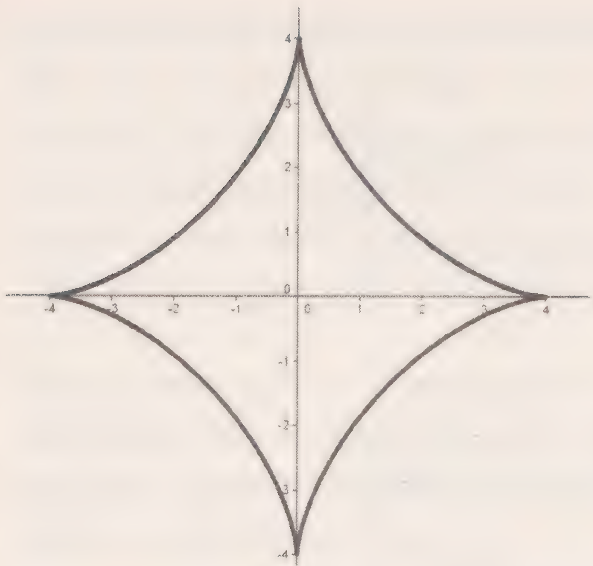
$$\begin{cases} x = r(K-1)\cos\theta + r\cos\big((K-1)\theta\big) \\ y = r(K-1)\sin\theta - r\sin\big((K-1)\theta\big) \end{cases}$$

As occurs in the epicycloids, if K is a natural number, the form of the curve has K unique points. A unique point is an 'angular' point in which a single tangent line at the point cannot be defined. If K is a rational number, it can be written as an irreducible fraction $K = p/q$, the curve is closed and has p unique points, while if K is an irrational number, such as $\sqrt{2}$, the curve is not closed and it occupies the whole space between the circumferences defined by R and $R-2r$.

If $K=1$, the hypocycloid is a point; if $K=2$, it becomes a line; if $K=3$, it is a deltoid with three unique points, as shown in the figure:

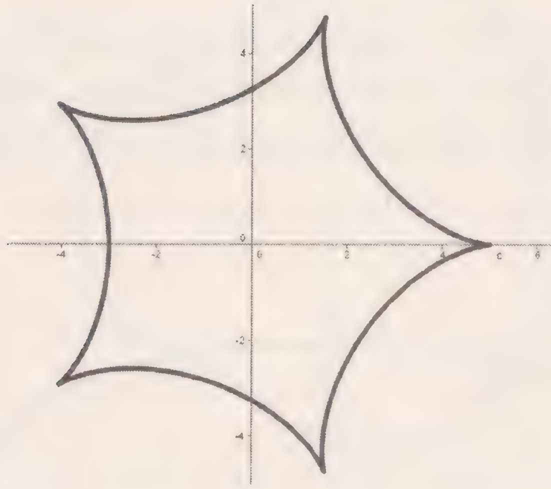


Deltoid.

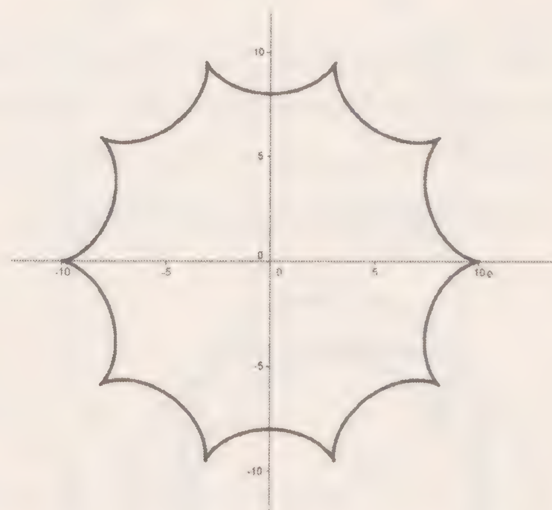


Astroid.

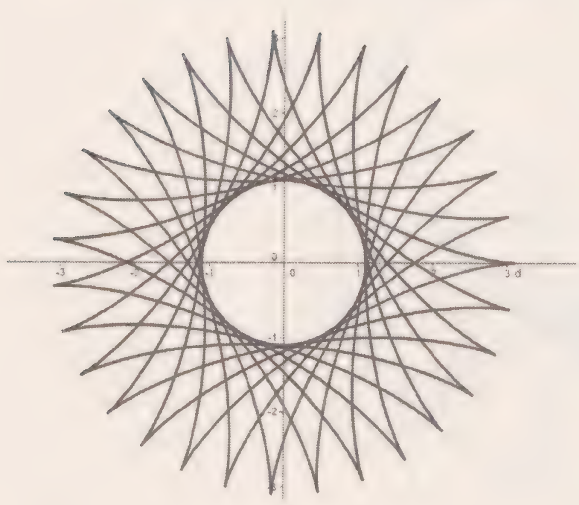
In the case that $K = 4$, the curve obtained is an astroid. It has four unique points and its equation can be written as a 'superellipse' due to the way that the equation reminds us of an ellipse. Its equation is $x^{2/3} + y^{2/3} = 1$. Its name alludes to its similarity with a star, although it is also known as a tetracuspid, cube cycloid or paracycle. Its form appears in the figure on the left. Some examples of hypocycloids for different values of K are:



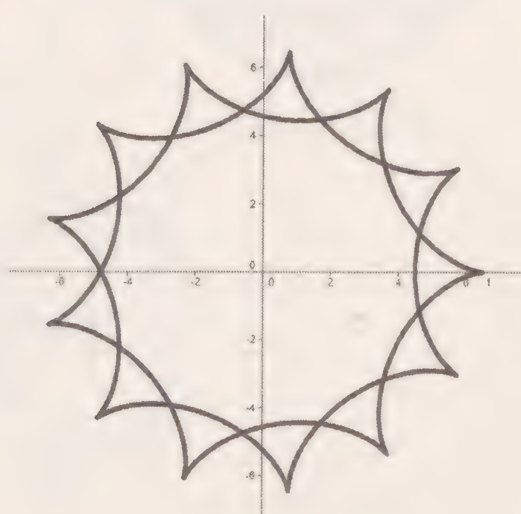
K=*5.*



K=*10.*



K=*31/10.*

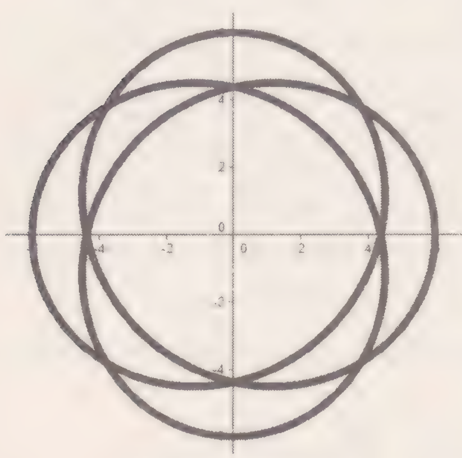


K=*13/2.*

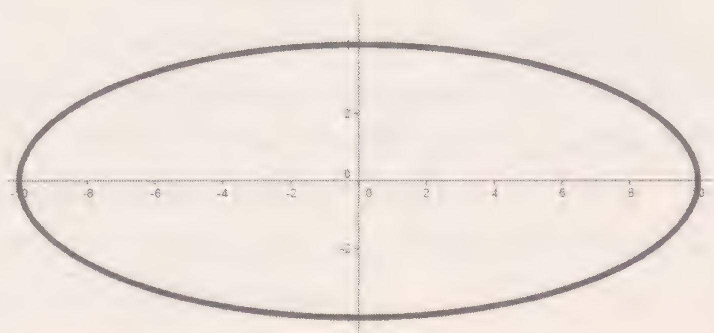
Another example of the curves generated by movement are the hypotrocoids and epitrocoids. In both cases they are curves of the trajectory from a point that is not located on the circumference but rather located at a certain distance from the centre of a circumference that spins on another, without dislodging itself. If the mobile circumference spins around the exterior part of the fixed one, the curves obtained are epitrocoids, whereas if it spins around the interior they are hypotrocoids. The polar equations of a hypotrocoid are

$$\begin{cases} x=(R-r)\cos\theta+d\cdot\cos\left(\dfrac{R-r}{r}\theta\right) \\ y=(R-r)\sin\theta-d\cdot\sin\left(\dfrac{R-r}{r}\theta\right) \end{cases}$$

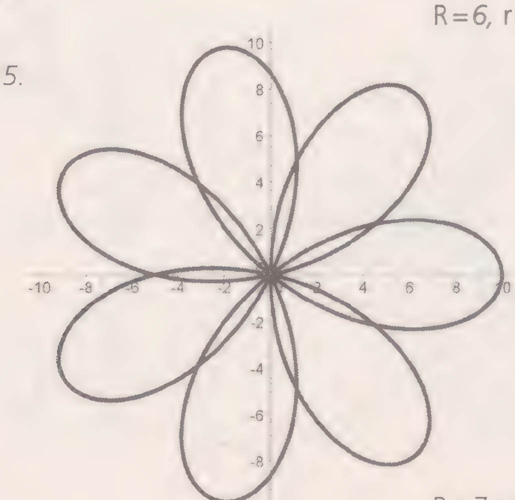
where R is the radius of the fixed circumference; r is the generatrix circumference that spins within the fixed one, and d is the distance from the point to the centre. When $d=r$ (the point is on the circumference), curves called hypocycloids are obtained. When $R=2r$, ellipses result.



R=4/3, r=1/3 d=5.



R=6, r=3, d=7.

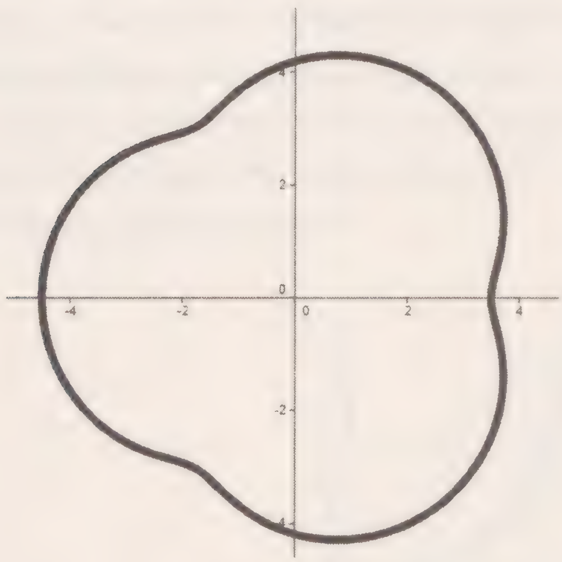


R=7, r=2, d=5.

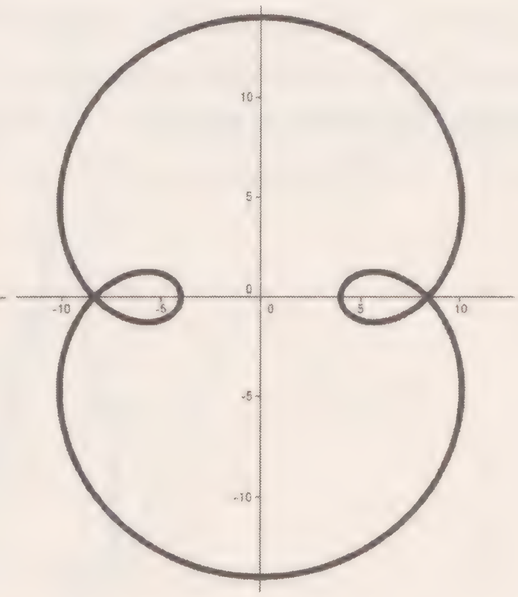
In the case of epitrocoids, the parametric equations are like this:

$$\begin{cases} x=(R+r)\cos\ \theta-d\cdot\cos\left(\dfrac{R+r}{r}\theta\right) \\ y=(R+r)\sin\ \theta-d\cdot\sin\left(\dfrac{R+r}{r}\theta\right) \end{cases}$$

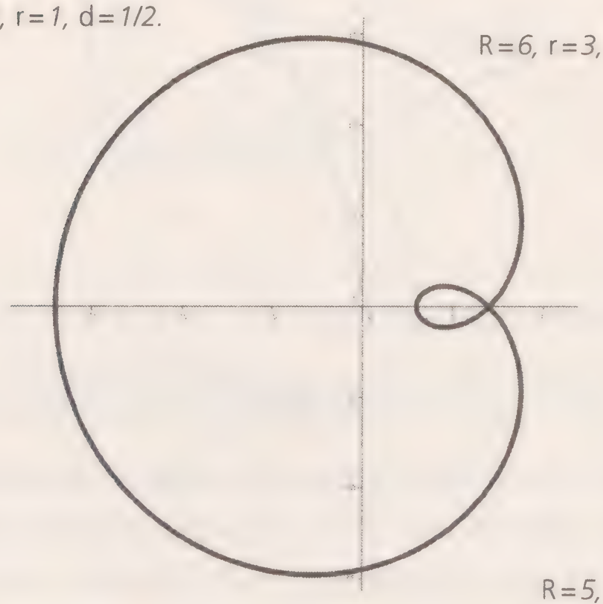
where R is the radius of the fixed circumference; r is the generatrix circumference that spins within the fixed one, and d is the distance from the point to the centre. Two special cases arise when $d=r$ (epicycloid curves result) and when $R=r$ (the snail or Pascal's *limaçon* occur).



R=*3*, r=*1*, d=*1/2*.



R=*6*, r=*3*, d=*5*.

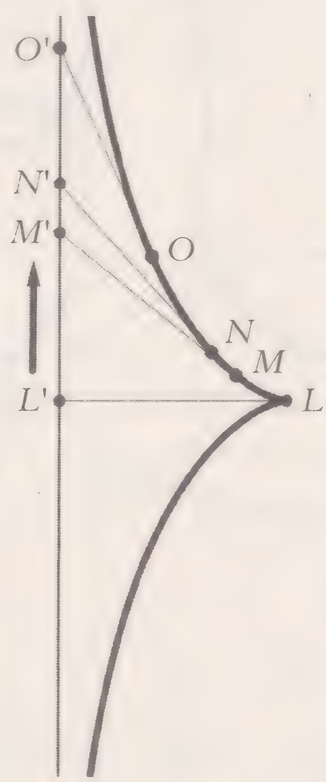


R=*5*, r=*5*, d=*7*.
Pascal's snail.

Pursuit or pulling curves

The pursuit curve is one that describes an object that moves at a constant speed w and which pursues another that moves in a straight line at speed v which is also constant. The pursuit curve is the movement which is made when it attempts to pursue someone in an optimum way. This, which a priori may appear very complicated to raise mathematically, is in reality a very simple and fun case, which enables us to know what can be done with basic mathematics.

The first thing needed to resolve the problem is to raise it adequately, and as such we need to understand how the pursuit is produced. The fundamental question is not to lose sight of the object that is pursued, go directly to it and correct the trajectory through which it moves. We start from the fact that both points move at a constant speed, but not necessarily the same one. From a mathematical point of view, the problem consists in drawing the line which connects the point that pursues with the point that is pursued.



Outline of the cinematic approach of the pursuit curve.

It was the French astronomer and mathematician Pierre Bouguer (1698-1758) who described this curve in 1732, as an example of what occurred when a dog followed his master, which is the reason it is also known as the dog curve. Later, Pierre-Louis Moreau de Maupertuis (1698-1759) generalised the problem of the

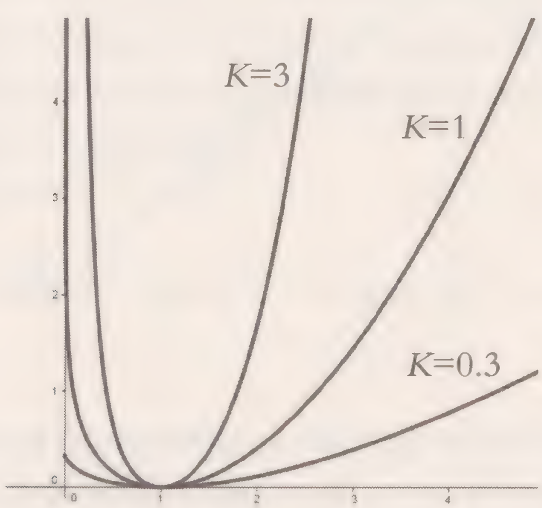
pursuit curve to other situations. Currently, the curve has applications in problems relating to robotics and the dynamics of tracking and surveillance.

As seen in other curves the relationship between two magnitudes defines an important part of the behaviour of a curve. It is defined as $K = \frac{v}{w}$, the relationship between the two speeds. The speed of the object that moves in a straight line is v, and the speed of the one that pursues it is w. As both speeds are constant, K is a constant. The equations are:

$$f(x) = \frac{1}{2}\left(\frac{1-x^{(1-K)}}{(1-K)} - \frac{1-x^{(1+K)}}{(1+K)}\right) \text{if } K \neq 1;$$

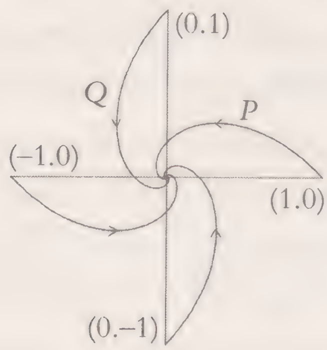
$$f(x) = \frac{1}{4}(x^2 - 2\ln(x) - 1) \text{ if } K = 1.$$

If $K = 1$, the two speeds are equal ($v = w$). If $K = 3$, one of the speeds is triple the other. According to the values of K, the forms of the curves are different, as can be seen in the following figure:

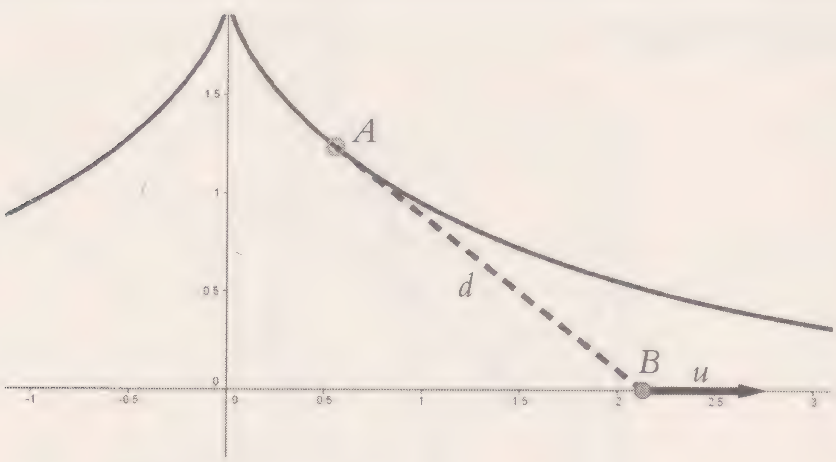


The process for writing the equations of these curves is quite complex and includes the calculation of the equation of the tangents, derivatives, minimising the calculations, solving differential equations ... The situation can be generalised to various types of movement and curves of vision. A classic example is the pursuit curve within a polygon. The approach to the problem is as follows: from each vertex of the polygon a dog or other animal that pursues another goes out so that the first never loses sight of the second. Similarly, the second does the same with a third animal, and so on successively until the last one does the same with

the first. Therefore, a chain pursuit is produced between the different points of each vertex. In the case of a square, the resulting curve is the one we can see in the following figure:



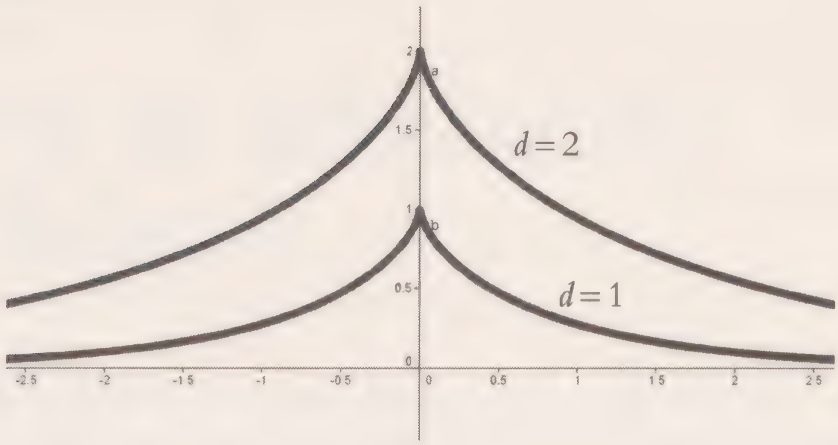
A different situation occurs when a pursuit is not produced, but rather a 'pull'. In this case a curve called tractrix is obtained.



The tractrix is a curve that describes an object (located at A in the figure) which is pulled by another (positioned at B), which is maintained at a constant distance d and which moves in a straight line. Its equation in polar coordinates can be written in the following way:

$$\begin{cases} x = d \cdot \left(\log \ (\tan \frac{\theta}{2}) + \cos\theta \right) \\ y = d \cdot \sin\theta \end{cases}$$

For different values of the distance d, the corresponding curves are:



This curve is known in the world of mathematics as the dog's bone curve. It is so named because of the following scenario. Imagine the master situated initially at the origin and the dog at *a*. The master would walk in the positive direction of axis *x*, while the dog, which would be pulled by the master's lead, would create resistance to return to point *a*, which is where the bone would be positioned.

This curve was studied by Claude Perrault in 1670 and subsequently by Newton and Huygens. One of its applications is the loudspeaker by P.G.A.H. Voigt, who patented it in 1927. It is known as the horn loudspeaker, and its design is based on the revolutionary surface of the tractrix, which is the surface that is generated when the figure rotates around one of its axes.

Chapter 4

Curves in Life, Society and Science

The presence of curves in everyday life, in society and in science is well known. This chapter will look at the most significant and frequently seen of these curves. For example, it is not unusual to assess the results of a medical scan, an ultrasound or an electrocardiogram from a curve, to determine the development of a child using percentile curves, to study the trends of population growth or look at the sound waves produced by a musical instrument.

Electric and magnetic curves

These curves are related to electricity, light and sound; they are waves in curve form. The mathematical forms of these curves belong to the family of sinusoidals, which are based on the form of the sine or cosine function. A relationship exists between the sine and cosine that allows it to pass from one to another with a simple change of variable:

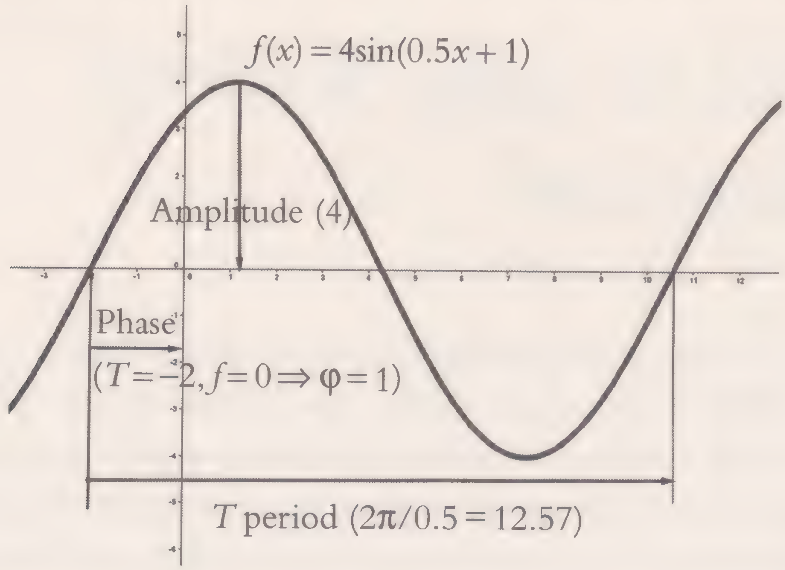
$$\cos x = \sin(x + \frac{\pi}{2}).$$

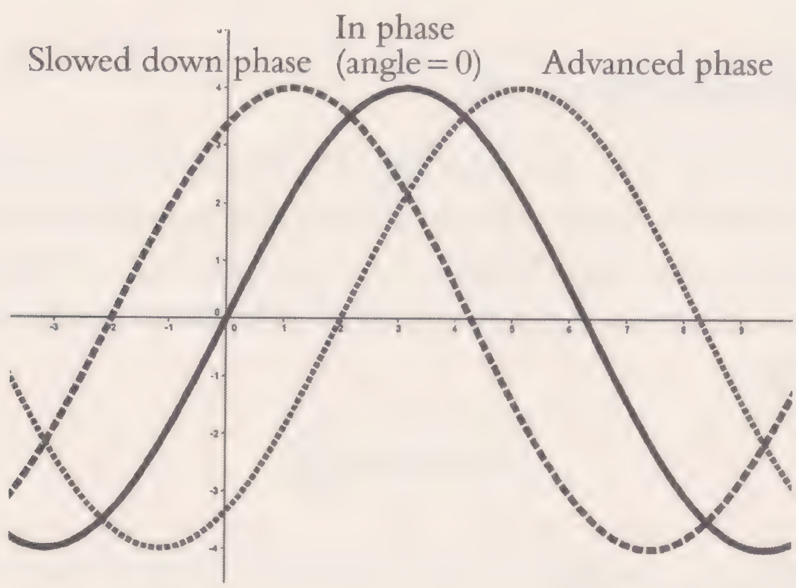
If we decide to only use the sine function, the sinusoidal functions can be defined from the waves in this manner:

$$f(x) = A \cdot \sin(\frac{2\pi}{T}x + \varphi) = A \cdot \sin(\omega \cdot x + \varphi),$$

where A is the amplitude of the function and indicates the maximum distance of the function from axis X. T is the period, which is no more than the time that it takes to complete a cycle, the time that the function takes in returning to the same state. φ is the initial phase and indicates the horizontal movement of a function; this value indicates if the function is ahead or delayed.

In the following figures we can see the details of the above definitions:





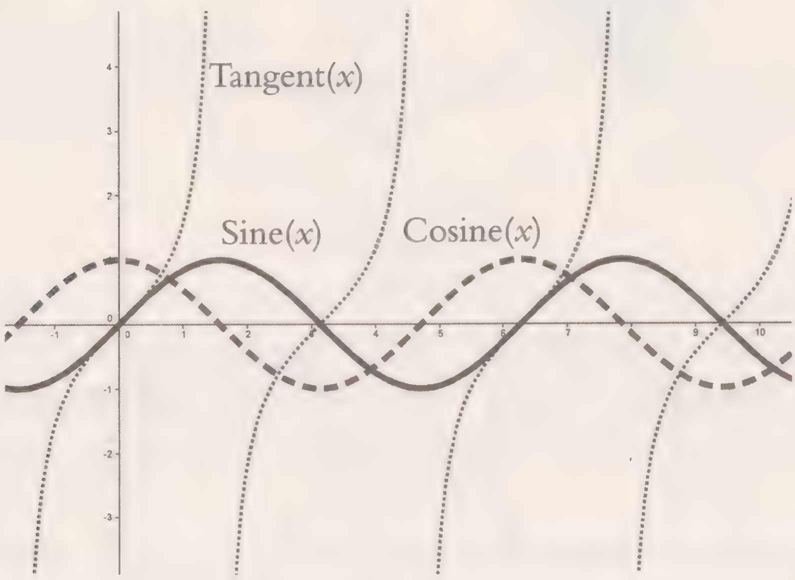
The alternating current appears as a sinusoidal function. The electricity flow constantly changes over time, increasing to reach its maximum and then decreasing to pass through zero before heading towards its next maximum. In this case, the variation of the magnitude and the direction is cyclical.

All modern electrical appliances that operate with alternating current have some relationship with the sinusoidal functions or with a combination of them. The basic trigonometric functions are sine, cosine and tangent. Between them relationships exist; the most well known are:

$$\sin^2(x) + \cos^2(x) = 1;$$

$$\tan(x) = \frac{\sin(x)}{\cos(x)}.$$

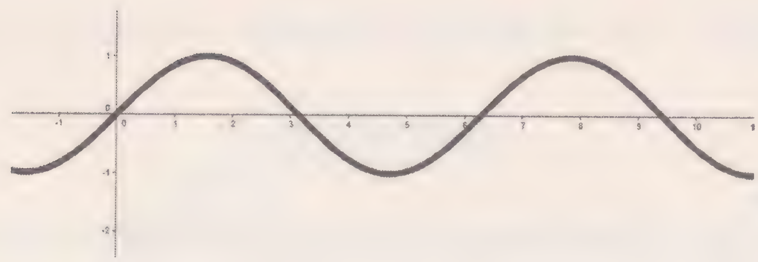
Therefore, to work with sinusoidal functions it is sufficient to know some of the functions, sine, cosine or tangent; the others can be deduced from the previous relationships. The graph representations can be seen in the following figure:



Other types of functions exist that also represent electrical curves. Some have the characteristic of being a periodic signal that shows some constant rates or surges and declines – these are *triangular waves*.

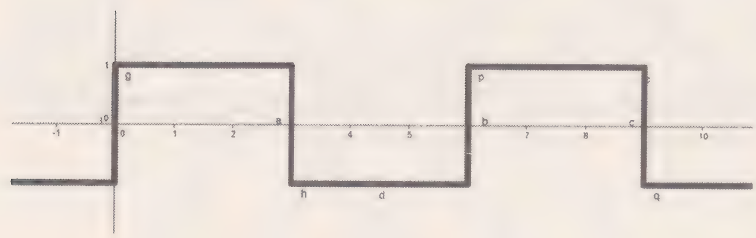
Others, such as *squared waves*, alter their values between two extreme limits without passing through intermediate values, in contrast to that which occurs with sinusoidal and triangular waves.

In other cases, the function passes through intermediate values, but on reaching the limit value it passes from the maximum to the minimum without passing through the intermediates. It can be said that the function takes the values between the maximum and minimum and on reaching an extreme it changes sign. This is the case of functions in the form of a *sawtooth*. These functions or waves are used to generate electrical impulses, with applications in physiotherapy, as they help to stimulate muscles; they are also used in digital electronics, to create images on television and computer screens, among other applications. In general, we have the image of a wave that oscillates gently – in other words, like a sinusoidal function. However, this is not always the case as can be seen in the following figures:

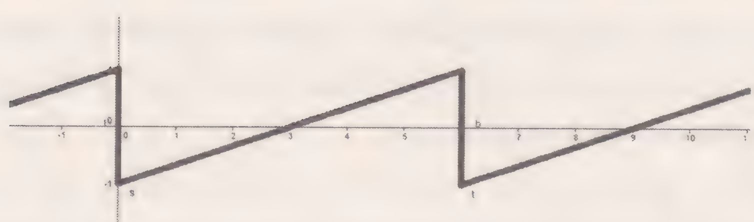


Sinusoidal wave.

Triangular wave.



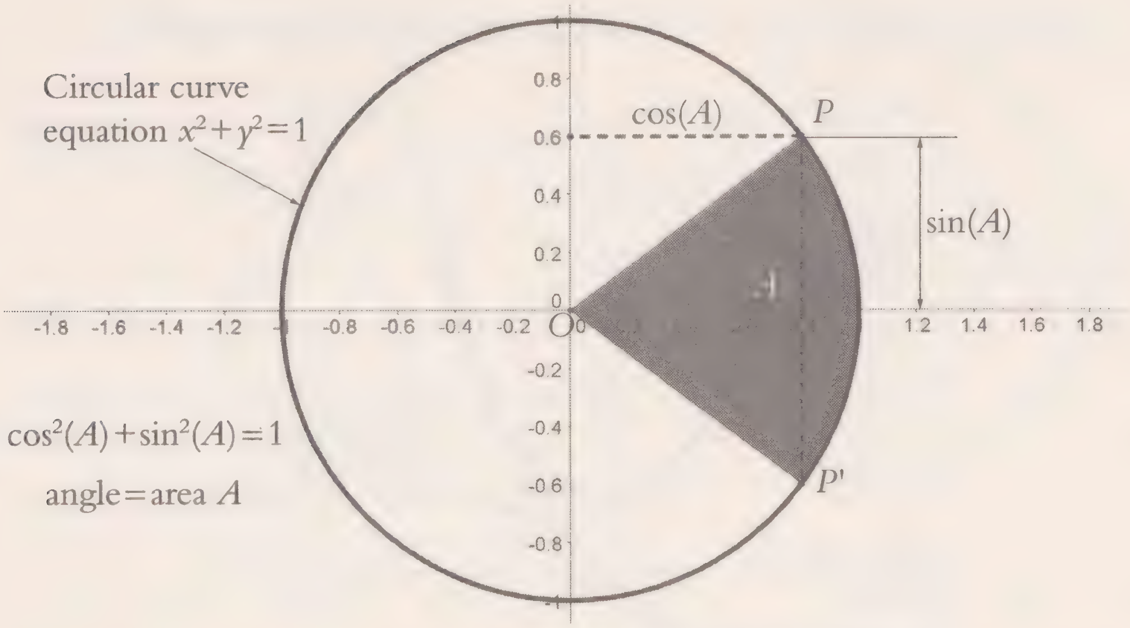
Square wave.



Sawtooth wave.

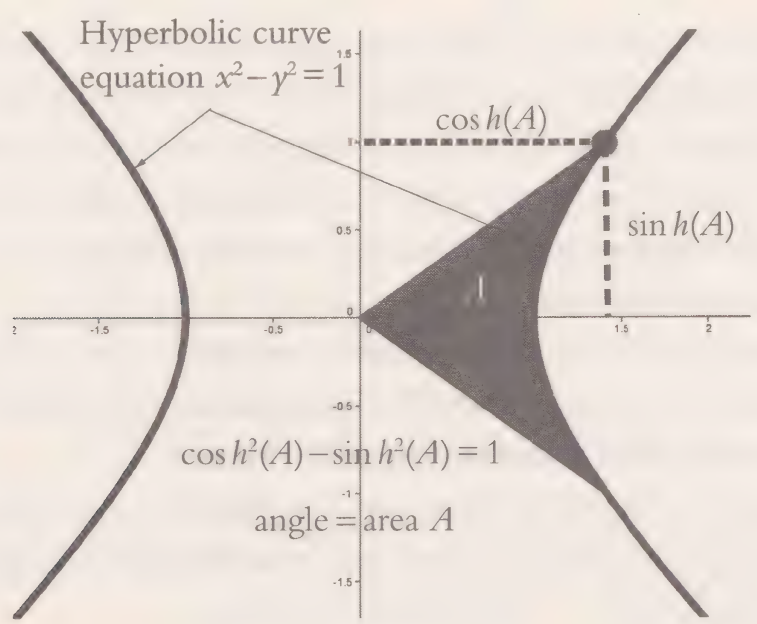
With regard to trigonometric functions, around the year 1760 Vincenzo Riccati (1707-1775) discovered *hyperbolic functions*. He discovered them when he was calculating the area under the hyperbola $x^2 - y^2 = 1$. Subsequently, Johann Heinrich Lambert (1728-1777) gave them the formulation which we know today.

While trigonometric functions are used to calculate the area under a circle, called circular functions, Riccati introduced the hyperbolic functions, to calculate the area under the hyperbola. These have the same geometric interpretation as circular functions.

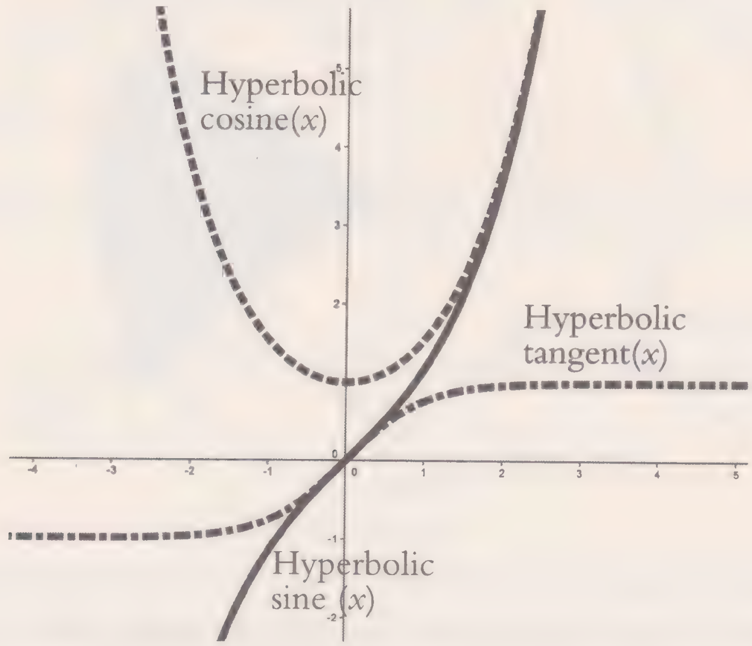


The analogy between the circular and the hyperbolic functions goes beyond the method for calculating the area and the relationships between them. The identities between the hyperbolae are parallel to those of the circulars. Every theorem or mathematical proposition in circular or trigonometric functions has a correspondent in hyperbolic functions.

The definition of these functions include the number e ($e=2.7182\ldots$) and they are combinations of exponential functions. Drawing a parallel with trigonometric functions, the value of the 'hyperbolic angle' A is the area that appears in the following figure. We can see the links with the trigonometric functions of the circle in the previous figure.



Hyperbolic functions have the following definition and representation:



$$\sin h\,(x) = \frac{e^{x} - e^{-x}}{2};$$

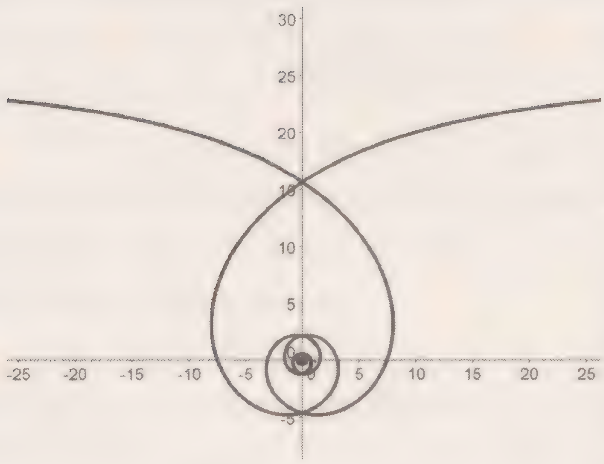
$$\cos h\,(x) = \frac{e^{x} + e^{-x}}{2};$$

$$\tan h\,(x) = \frac{e^{2x} - 1}{e^{2x} + 1}.$$

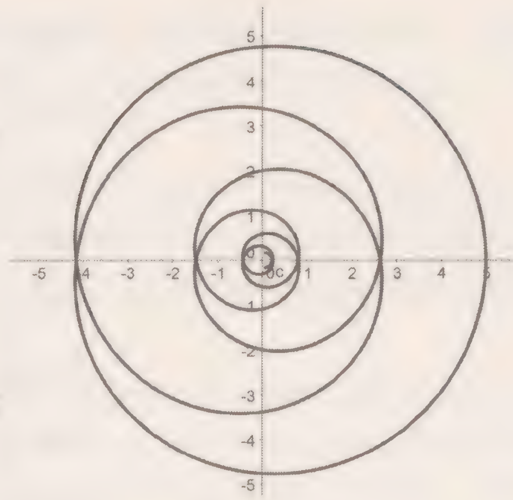
The importance of hyperbolic functions lies in the fact that they appear in the solutions of determined mathematical problems. Their presence and relationship with situations in daily life, along with scientific problems, is considerable. They are found in the solution to some problems of electromagnetic theory, in the transfer of heat, fluid dynamics, in special relativity, in Poinsot's spirals, in catenaries, in mechanics, in logistic functions...

The hyperbolic cosine function represents a catenary curve, as can be seen in the previous figure. It is a curve that is formed by the effect of gravity when freely suspending a rope or chain from two extreme points.

Trigonometric and hyperbolic functions also maintain a relationship with other special curves. This is the case of Poinsot's spirals, which are formed by representing the hyperbolic relationships of sine and cosine as a graph.

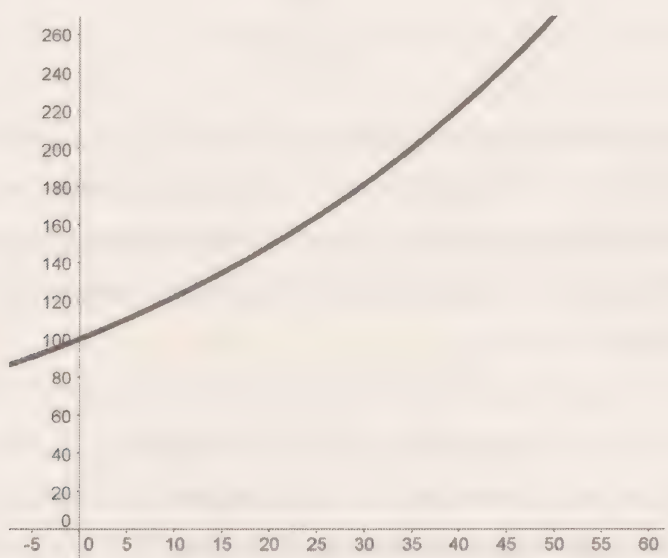


$$r = \frac{5}{\sinh(\frac{\theta}{5})}.$$

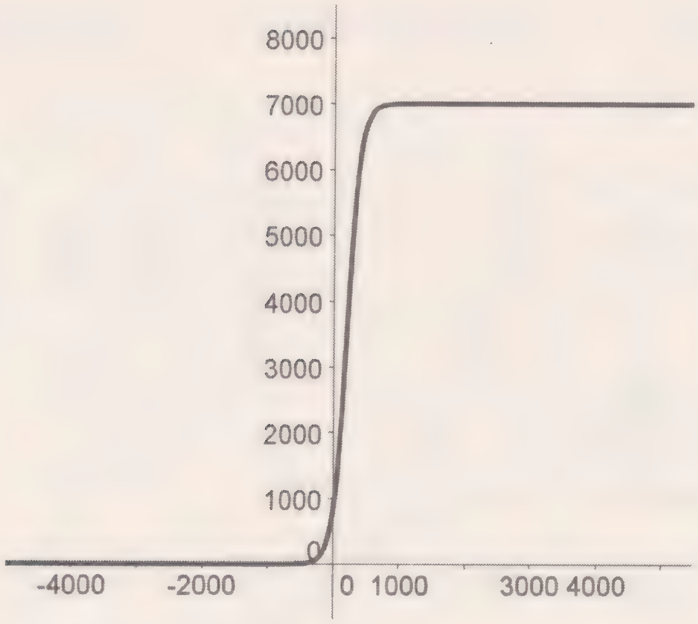


$$r = \frac{5}{\cosh(\frac{\theta}{5})}.$$

Other related curves with hyperbolic functions are *logistical curves*. Pierre François Verhulst (1804–1849) studied them around 1845 in relation to population growth. Their application to scientific processes, research, social development and day-to-day actions is very important. These are curves that model processes of change with time, from low levels at the start, approaching maximum values after a certain time has passed. The change from a low to a high level is produced quickly. Its graph looks like an 'S'.



Curve of exponential growth of a population. Cartesian equation: $P(t) = 100(1 + 0.02)^t$*,*
where $P(t)$ *= population at time* t*. Initial population (time 0) = 100;*
growth rate = 0.02 (2%).



Logistical curve of a population. Cartesian equation: $P(t) = \dfrac{7000}{1+e^{2-0{,}01t}}$ *for a maximum population of 7,000 individuals.*

Its simplest expression is:

$$f(t) = \frac{1}{1+e^{-t}},$$

and in general it has:

$$f(t) = \frac{K}{1+a \cdot e^{-b \cdot t}},$$

where K, a and b are constant parameters.

This function represents the behaviour of certain natural systems that grow exponentially, and at the end of a given time a rival element appears among some of its members that reduces the growth until it ends up stopping. It has applications in economics, biology, medicine…

The relationship of logistical curves with artificial neuronal networks is noteworthy in the study of neuronal relationships and the development of artificial intelligence. It also has applications in sociology, economics, political science, psychology, in the description of learning processes, marketing (for studying the introduction of new products onto the market) and so on.

Let us now consider some specific examples. The embryonic growth of a fertilized egg makes cells begin to divide and to grow $1, 2, 4, 8\ldots$ times, doubling their number

each time. This growth cannot be unlimited, given that the uterus cannot grow beyond a certain size. Therefore, from a certain moment growth reduces up to the birth of the foetus.

The same occurs in the case of population growth, whose increase depends on the number of existing people and resources available. If the population grows, the resources will reduce and, therefore, growth will stop increasing until it stabilises according to the resources.

Other applications are found in the field of medicine and, more specifically, the evaluation of the growth of tumours. The size of the tumour is determined, and then the rate of its growth and the effect of chemotherapy to eliminate cancerous cells is considered. From this information, the equation of the curve that will indicate the development of the patient's tumour is calculated. Equally, they are applied to the spread of epidemics and diseases such as seasonal flu. Percentages of people infected are predicted for the purpose of distributing vaccines and other resources to best assist all patients.

Lissajous or Bowditch curves

These types of curves were studied by Nathaniel Bowditch (1773-1838) in 1815 and, more extensively, by Jules Antoine Lissajous (1822-1880) in 1857. Lissajous curves have applications in astronomy, physics and medicine. They are constructed from two harmonious oscillations, combining them in a perpendicular manner. It is like putting two waves in contact, two curves, with some particular characteristics. The waves are harmonious and, therefore, they have a movement that repeats in time (it is said to be periodic) and is defined by a sine or cosine function. These curves can be observed using an oscilloscope.

The patterns of the Lissajous curves are produced within a rectangular limit framework. The parametric equations of this family of curves is:

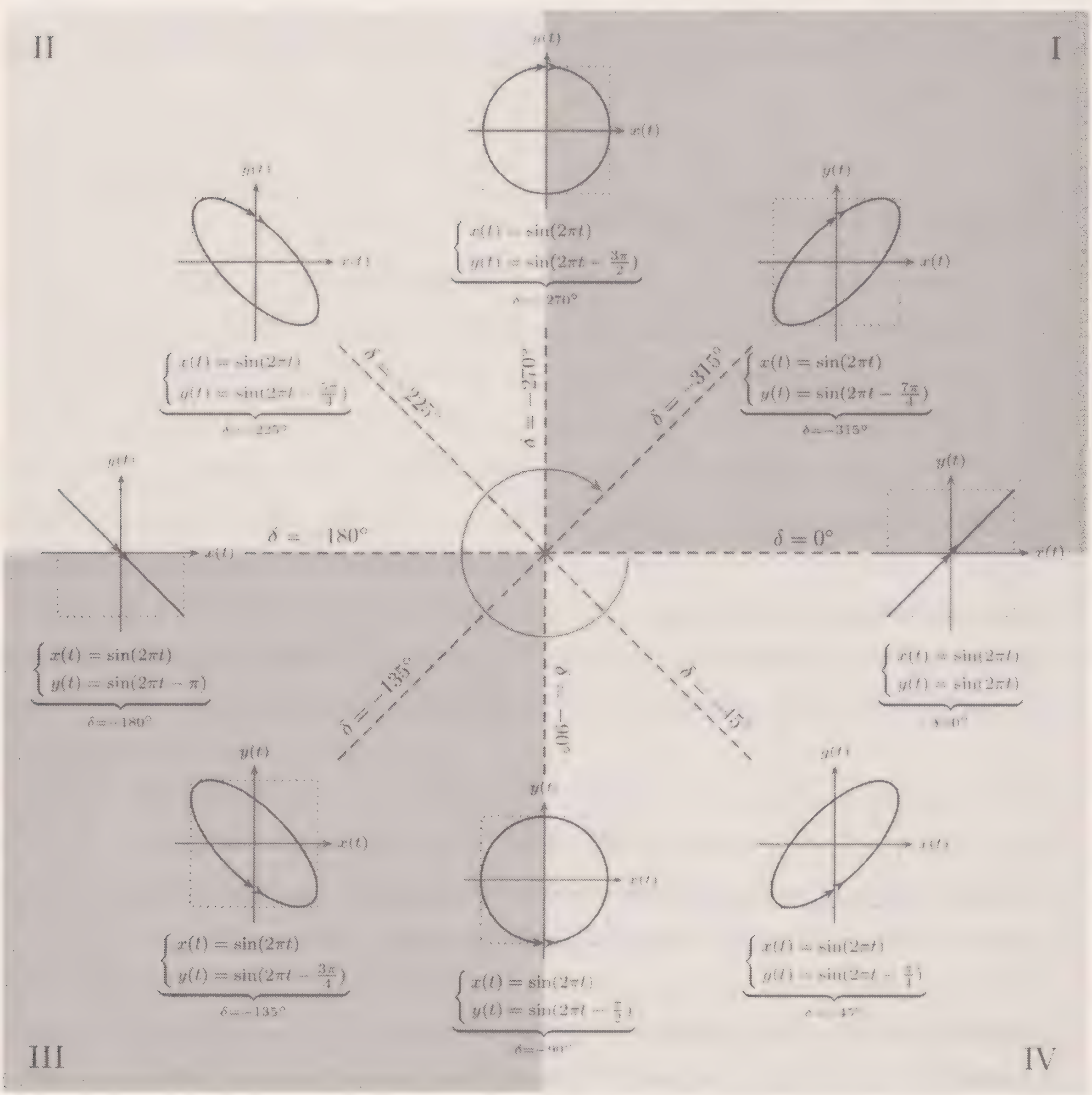
$$\begin{cases} x = A \sin(a \cdot t + \delta) \\ y = B \sin(b \cdot t) \end{cases}$$

In these equations values that are typical of waves appear, such as amplitudes (A and B), frequencies (a and b) and the phase lag between waves (δ).

This type of curve is very sensitive to the relationship between the frequencies of waves $\frac{a}{b}$ and the phase lag δ. According to the value of the phase lag δ and the relationship between the amplitudes A and B, the curve obtained varies between a

line, a circle and an ellipse with different inclines, as indicated in the table and the lower figure.

a/b	*A* and *B*	Lag (δ)	Note
1	Any	Any	Ellipse
1	$A=B$	$\delta=\frac{\pi}{2} rad$	Circle
1	$A=B$	$\delta=0$	Line
2	Any	$\delta=\frac{\pi}{2} rad$	Parabola

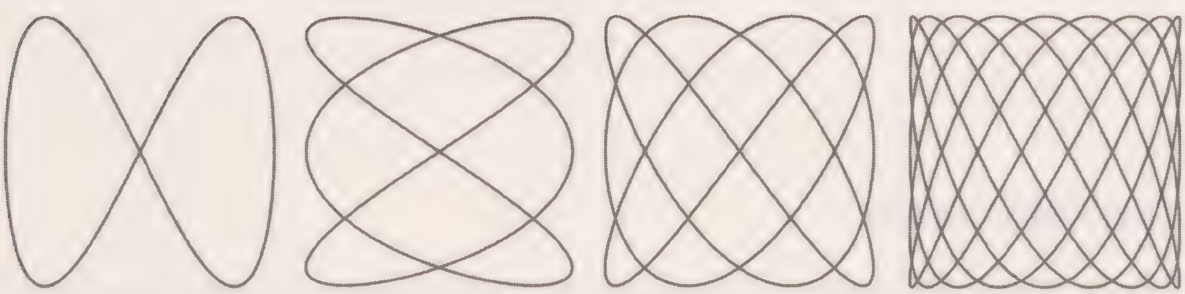


In the case that the relationship $\frac{a}{b}$ is a rational number, the curve is closed; if it is an irrational number, for example $\sqrt{2}$, the curve is not closed and it occupies all the space. This space is framed by amplitudes A and B. These define the rectangle in which the Lissajous curves are 'enclosed': if A and B are equal, the rectangle is converted into a square; if they are different, the figures are framed in rectangles. They have the same appearance, but are deformed by changing the square to a rectangle.

If the curves are drawn

$$\begin{cases} x = A\ \sin(a \cdot t + \delta) \\ y = B\ \sin(b \cdot t) \end{cases}$$

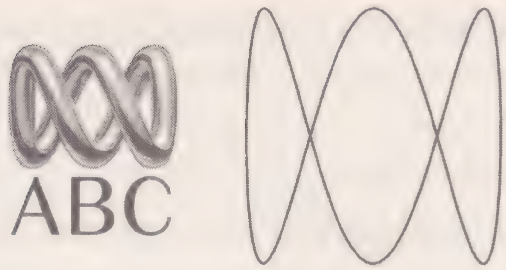
for different values of A, B and a, b with $\delta = \frac{\pi}{2}$ rad, the following result is obtained:



A=1, B=1 a=1, b=2 | A=1, B=1 a=3, b=2 | A=1, B=1 a=3, b=4 | A=1, B=1 a=5, b=9

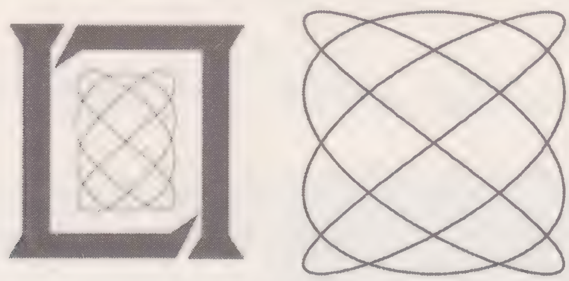
A=1, B=1 a=5, b=5 | A=2, B=1 a=1, b=1 | A=1, B=1 a=2, b=1 | A=1, B=1 a=$\sqrt{2}$, b=1

These curves are used in graphic design, and some companies have incorporated them into their logos, for example the Australian Broadcasting Corporation:



$A=1$, $B=1$, $a=1$, $b=3$ *and* $\delta=\frac{\pi}{2}$ *rad.*

The Lincoln Laboratory also has an image with a Lissajous curve:



$A=1$, $B=1$, $a=4$, $b=3$ *and* $\delta=0$.

Sonorous curves

Let us consider the circular trigonometric functions seen at the start of the chapter, but with a small limitation: the waves are formed by the vibration of a curve that is fixed at both ends. In this situation, the approximation of a harmonic oscillator is produced; this is a vibration system that, if released, returns to a position of equilibrium, describing damped waves or sinusoidal curves (which gradually reduce with time due to friction). A graphic example of a harmonic oscillator is that of a spring with a hanging mass. If the mass is moved, it returns to a position of equilibrium with a movement in the form of a circular trigonometric function. As before, the equation for the curve can be written as

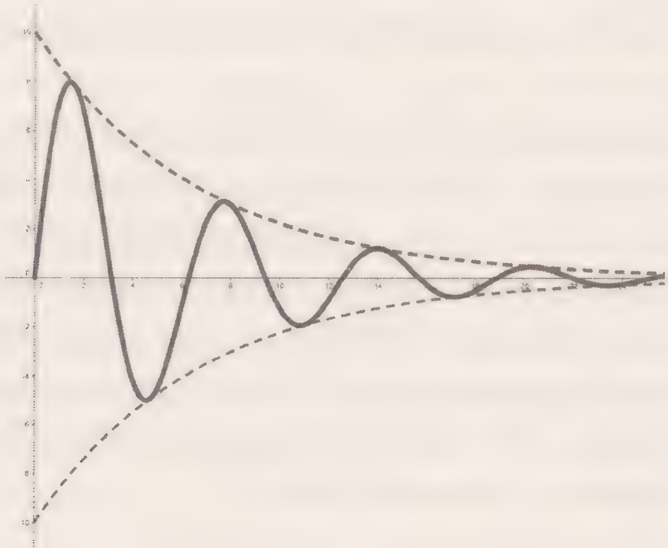
$$f(t)=A\sin(\omega t+\varphi).$$

It can be seen, in contrast to the previous case, that the equation does not correlate the position with time. The corresponding representation is the sine function. In the case that the oscillation is damped by friction, the result shall be a sinusoidal function modelled by an exponential of damping. In other words, the trigonometric

function becomes increasingly smaller. In this case the equation that rules the movement is:

$$f(x)=A\cdot e^{-kt}\sin(\omega t+\varphi).$$

It is the same as the previous one, but with a damping factor that is written mathematically with the number e raised to a negative value. In the equation, k depends on the initial conditions. If we take the values of $A=10$, $k=0.125$, $\omega=1$ and $\varphi=0$, the following graph results:



Graph representation of a damped harmonic oscillatory movement.

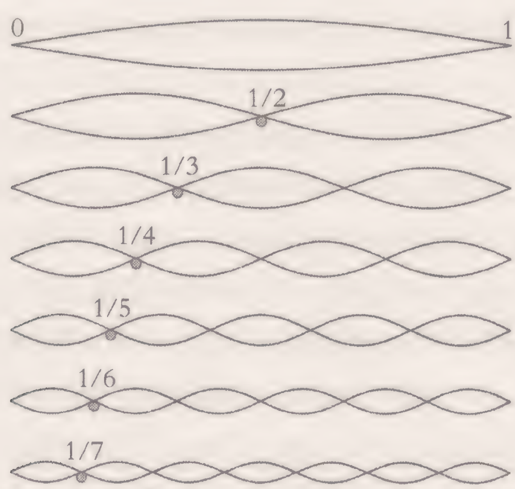
With discontinuous lines the value of the exponential has been marked $g(x)=10e^{-0.125x}$, which is simply the decrease of the oscillation due to friction. Ultimately, we have the graph of the sine function which reduces exponentially. The element that represents the curve is the loss of energy that is produced through friction, and in the case of a sound wave the limit case is silence.

Another possibility is to consider an oscillatory movement not as a product of releasing the system that makes the movement, but rather by applying a force to it that varies with time in a sinusoidal form. The result is a movement with forced oscillation that will create a stationary oscillation, in other words, that will not change with time. One example would be a swing in the park. In this case, the oscillatory movement is produced by the swinging of a child on a swing. If it moves from the initial position, the swing will oscillate before stopping again at the initial position – it is a damped oscillatory movement. In the case that someone helps the

child by pushing them and keeping the swing oscillating, we shall have a forced oscillatory movement that makes the swing move constantly with time, in other words, stationary. Only if the movement of pushing coincides with the oscillation of the swing can the amplitude of the movement increase with a small added force. If the pushing doesn't coincide, the movement will vary significantly.

This can represent the *resonance*, the set of phenomena related to periodic movements in which the reinforcement of the oscillation is produced by other oscillations. The resonance can be observed in mechanical movement, in music and in electrical circuits; in the case of magnetism it has applications in the field of medicine with magnetic resonances as diagnostic methods, it also serves as an explanation of nuclear motion.

All these oscillatory phenomena are clearly seen in acoustic oscillations. In the case of sound, the resonance of waves is very significant. Oscillation is produced at different frequencies and, as such, stationary waves are created. These resonance frequencies are integral or harmonic multiples. In musical instruments an oscillatory movement is produced due to the vibration of a string, a membrane or the air. To visualise the situation an example is shown in the following figure. The string has fixed points at each end, called nodes. Each harmonic is divided into 2, 3, 4… parts. The oscillation frequency is increasingly higher:



On dividing the string into two parts it has double the frequency and, therefore, it is an octave above the first note. If it is divided into three parts, we get a fifth; with four parts, a fourth, and so forth. It can be said that each sound is formed through a

combination of many simple oscillations. In music, the characteristics, the forms of reinforcing some sound sensations, are studied in musical harmony.

In the West, the scale of notes used is the chromatic scale, which is based on octaves and frequencies of a harmonic series, being integral multiples of the fundamental frequency; they are related to each other by relationships of integral numbers and small proportions of the whole number.

Our ears tend to group related frequencies into harmonies. If a sound is heard that is composed of only a few simultaneous tones, and if the intervals between the tones form part of a harmonic series, our brain tends to group this input into a sensation of the whole series, although the fundamental tone is not even playing. This phenomenon is used in recorded music, especially with low tones, which cannot be reproduced in small speakers. As such, introducing part of the corresponding harmonic sounds offsets the limitations of the musical equipment.

Returning to resonance, this phenomenon is produced, in general, when a body can vibrate and it is subject to a force at constant intervals. These intervals coincide with the vibration period characteristic of this body, and its effect is to increase the amplitude of oscillation with a relatively small force. In these circumstances the body vibrates, increasing the amplitude of the movement after the addition of each successive addition of force.

This effect can be destructive in some rigid materials, such as a glass that breaks when a soprano sings, sustaining the same resonance frequency. For the same reason marching soldiers break their stride when they go over a bridge because the frequency of all their steps together can coincide with the resonant frequency of the bridge. The old Tacoma Narrows bridge was a suspension bridge 1,600 m in length, the third longest in the world at the time (1937) and it was situated in the U.S. state of Washington. The bridge collapsed due to the resonance produced by the complex oscillation of the wind as it passed between the girders.

When a radio is tuned the internal circuit of the device is set to operate at a natural frequency that resonates with the emission frequency of the transmitter.

The buildings erected in areas with a probability of earthquakes are built to withstand the frequencies of ground movement. This means that they are more stable with small oscillations. Seismic waves are oscillations due to the forces that are generated in the Earth's interior and which spread across the surface.

Internal seismic waves are divided into two groups: primaries (P) and secondaries (S). P waves are generated by the compression and expansion of the ground in the direction in which it is spreading. In S waves the displacement is transver-

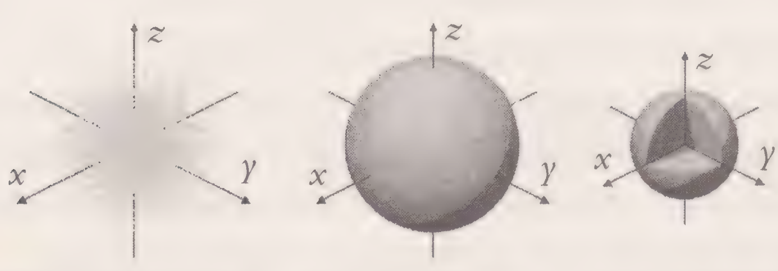
sal to the direction of the spread. Their speed is slower than the primary waves and for that reason they appear at ground level some time after the P waves. The S waves are responsible for generating surface vibrations and thus produce the majority of damage.

Diffusion curves: zones of movement

So far curves have been presented according to whether they described some movement or if they were generated by movement. There are some situations in which the movement that is produced does not follow a trajectory like a curve, but rather it is produced within a specific area, with a determined probability in a determined area in space. A well-known example is the orbitals of electrons within an atomic nucleus.

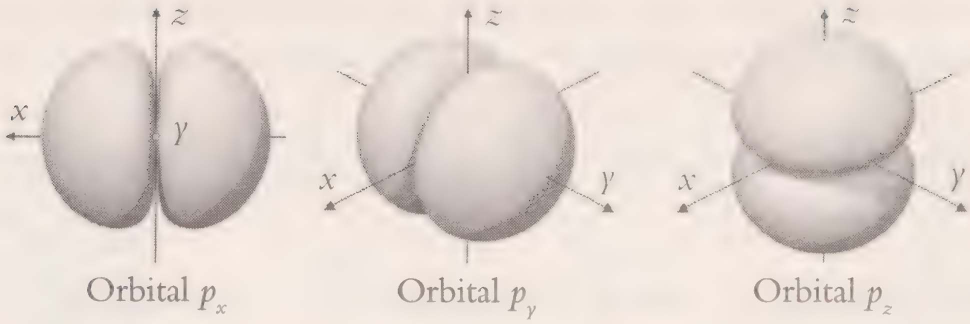
The concept of the atom is fundamental to chemistry. They are composed of electrons, neutrons and protons. The atomic structure is the specific form of an atom. The nucleus has protons and neutrons, whereas there is a cloud of electrons surrounding it. Depending on the atom, the trajectories of the electrons have different forms.

The area in which an electron moves is called an orbital. Each type of orbital has four associated quantum numbers. The first defines its size (n), and the second its geometric form (l): When $l = 0$, this is an orbital s (sharp); if $l = 1$, it is the orbital p (principal); if $l = 2$, it is the orbital d (diffuse), and if $l = 3$, it is the orbital f (fundamental). The third quantum number is the magnetic moment (ml), which determines the spacial orientation of the orbital, and the fourth is called spin. The following figures show the 3D forms that determine these 'zones of movement':

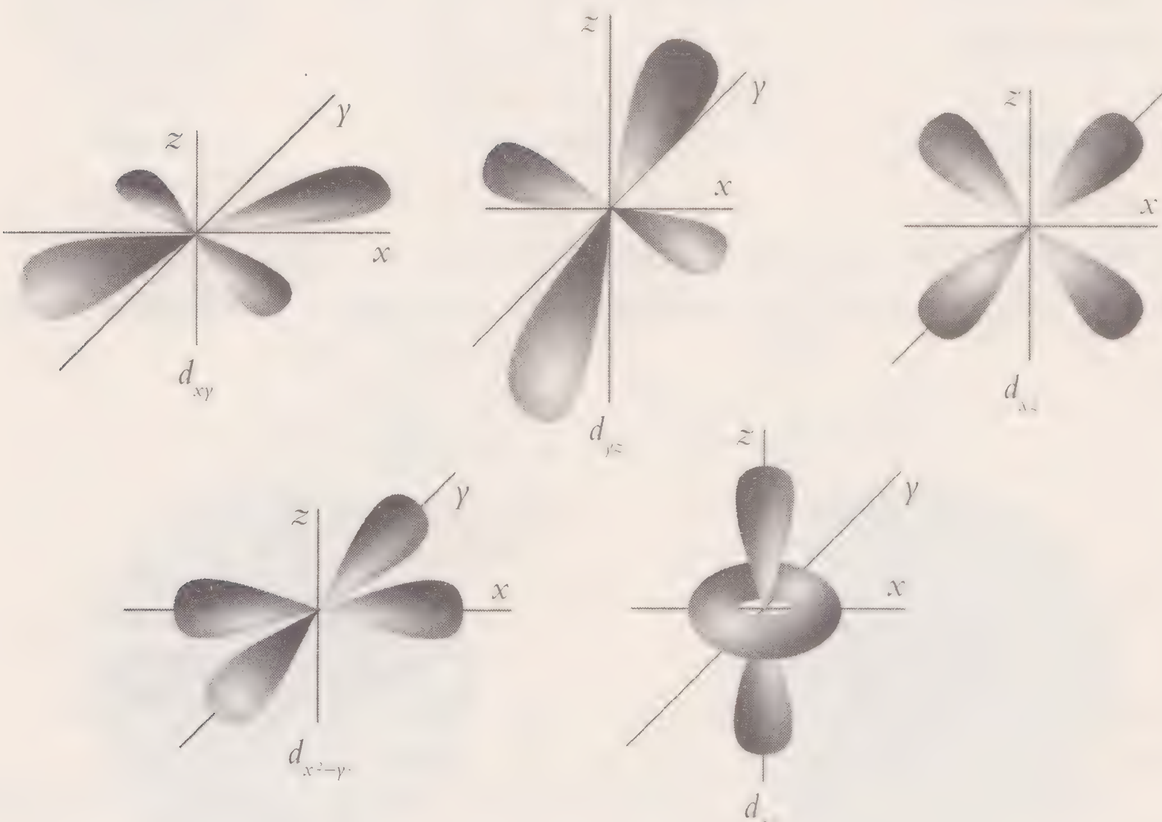


Orbital s ($l = 0$, $m_l = 0$) $n = 1$ (hydrogen) $n = 3$ (sodium)

Orbitals s (l = 0) spherical in form.



Orbitals p *(*l*=1) with two lobes.*

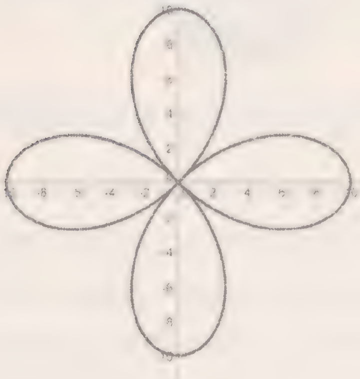


Orbitals d *(*l*=2) with four lobes. There are five types:* d_{xy} d_{yz} d_{xz} *(on planes* xy, yz, xz*);* $d_{x^2-y^2}$ *(hyperbola on plane* xy*), and* d_{z^2} *on axis* z.

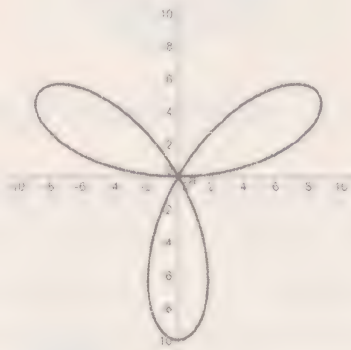
Observed mathematically, these 3D figures recall the different curves of the family of *rhodonea curves* or rose curves. These curves were studied by Luigi Guido Grandi between 1723 and 1728. The rose curve or rhodonea curve is a sinusoidal curve with this equation in polar coordinates:

$$r = a \cdot \cos(k \cdot \theta) \text{ or } r = a \cdot \sin(k \cdot \theta).$$

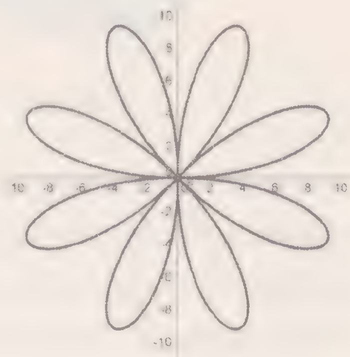
If k is an integral, the curve has $2k$ petals if k is even and k petals if it is odd. In both cases, the curve is closed. If $k=1$, the rose is a circle.



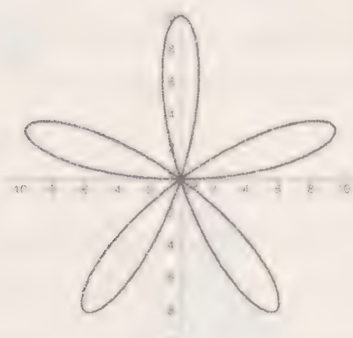
$k=2$
$r=10\cos(2\theta)$



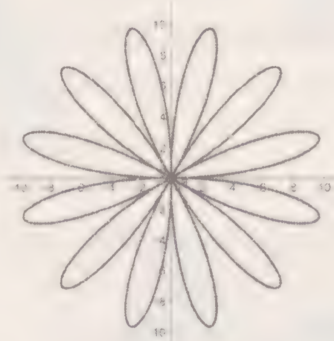
$k=3$
$r=10\sin(3\theta)$



$k=4$
$r=10\sin(4\theta)$



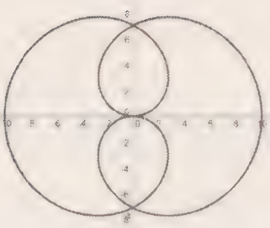
$k=5$
$r=10\sin(5\theta)$



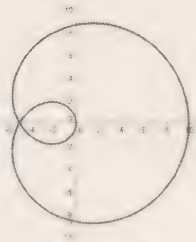
$k=6$
$r=10\sin(6\theta)$

When k is even, the drawing of the entire rose is completed varying the angle θ between 0 and 2π. If k is odd, the drawing of the entire rose is completed by varying the angle θ between 0 and π.

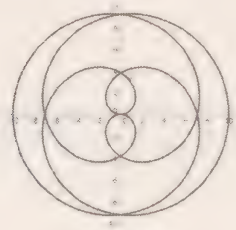
If k is rational and is written in the form of a fraction, the curve is also closed with a finite length. With $k=\frac{n}{d}$, the form of the curves depends on the values of n and d. We can predict the number and form of the petals according to their values. The whole drawing of the rose will be completed by varying the angle θ between 0 and $\pi \cdot d \cdot p$, where $p=1$ if $n \cdot d$ is odd, and $p=2$ if $n \cdot d$ is even. Here are some examples:



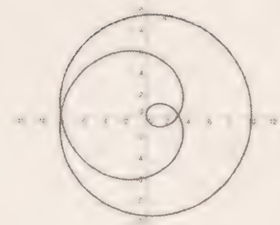
k = *1/2*



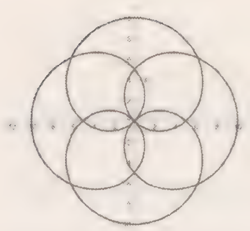
k = *1/3*



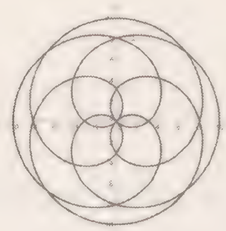
k = *1/4*



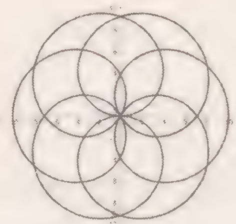
k = *1/5*



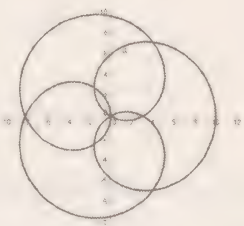
k = *2/3*



k = *2/5*

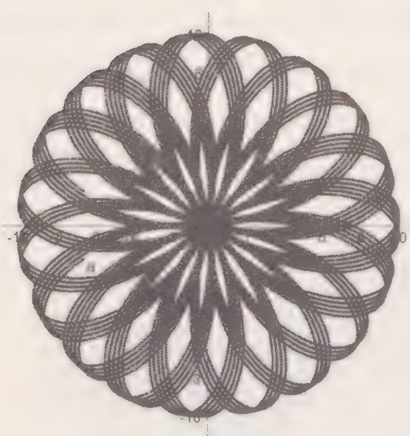


k = *3/4*

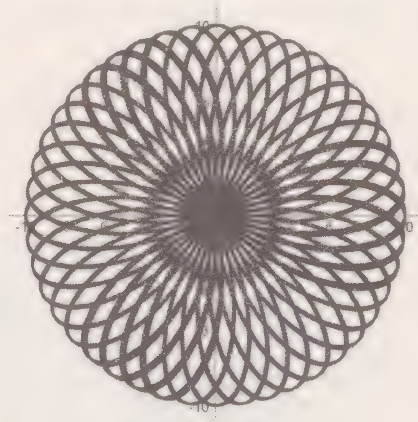


k = *3/5*

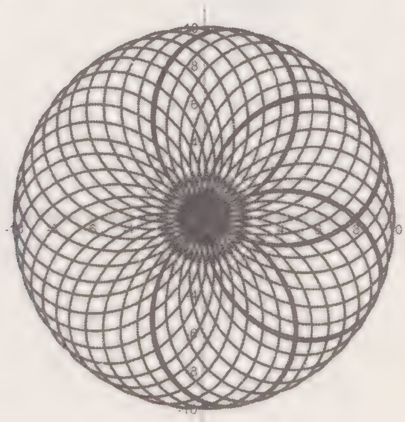
If the number k is irrational (for example, $\sqrt{2}$ or π), the rose occupies all the space, creating a dense figure, with an infinite number of petals:



k = e



k = π



k = $\sqrt{2}$

Chapter 5

Curves in Nature, Art and Design

Conic sections

The fundamental importance of conical curves lies in the operation of the human visual apparatus, the capacity of perception of which depends principally on the eye and the light rays that reach it. The field of vision within its scope forms a cone, in which a biconvex lens interjects as a lens. All images of the optical reality, the entire perspective, the whole projection, are presented in the form of a conic section. Therefore, it is no exaggeration to describe our world as a 'world of conic sections'.

From the development of perspective in the Renaissance onwards, the concepts of conics have been central to the evolution of art. Conical curves are so called because, mathematically, they can be obtained as sections (slices) through a body called a 'two-sheeted cone' viewed on a plane.

There are three distinct ways of cutting a two-sheeted cone; each one will generate a different conical curve.

Parabola: parallel section to the generatrix of the cone.

Circle and ellipse: sections that cut the generatrix of the cone.

Hyperbolae: parallel section to the axis of the cone.

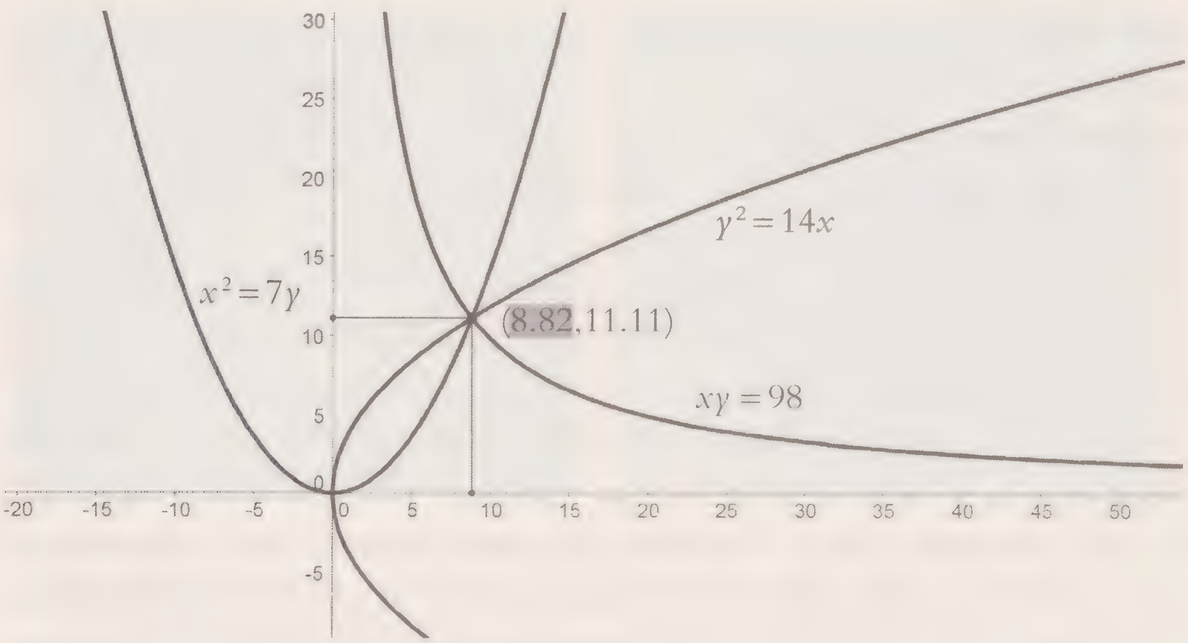
The first figure demonstrates the conic section obtained by means of a parallel plane to its generatrix; the section is a parabola. The second figure shows two slices, or sections, that cut the cone axis and produce a circle when the cut is perpendicular to the axis and an ellipse when it is oblique. Finally, in the third figure, the section is obtained using a plane parallel to the axis of the cone. The curve obtained consists of two parts or arms, one on each side of the cone; it is a hyperbola.

Menaechmus (c. 380–c. 320 BC) was a Greek mathematician from the Roman province of Tracia (located today between Bulgaria, Greece and Turkey). He was the first to work on conic sections to try to resolve one of the three famous problems from classical antiquity, the so-called Delian problems. It appears that he was Alexander the Great's professor, and from their relationship there is an interesting anecdote: Alexander asked him for an easy way to understand geometry and Menaechmus responded: "Oh King, to travel from one place to another there are paths for the king and paths for the people, but in geometry there is only one path for everyone."

The first Delian problem was raised in the year 370 BC. The oracle asked the inhabitants of Delos what length the side of a new cubic altar should be to give double the volume of the one that existed at that time, which had a 7-metre edge. The philosopher and mathematician Plato advised the Delians to forget the magical interpretations of the oracle and concentrate on solving this problem mathematically. Menaechmus, one of the most brilliant pupils of Eudoxus from Plato's Academy, was the one to solve this problem. To do so he had to elaborate on an initial version of the curves that are produced in the conic sections. In particular, the parabola and hyperbola. He worked with a particular cone (straight, or in other words a 90° angle at the vertex), which was cut by a plane perpendicular to its generatrix. He made a series of calculations and constructions of this section, which was a parabolic curve. Later he repeated this work on similar sections in another two types of non-straight cones (acute and obtuse).

In the search for the form of these curves, Menaechmus undoubtedly used mechanical contraptions, mathematical tools such as proportionality and some primitive algebraic and representation methods. For that reason it is thought nowadays that the works of Menaechmus and the eight books of the *Conics* by Apollonius of Perga constituted a primitive form of algebra and coordinates systems. They were translated from Greek to Arabic by mediaeval mathematicians such as the Baghdadi Thabit ibn Qurra (9th century) and collated by the Vatican Library in 1536, and were the historic bases of the great findings of Descartes and Fermat.

With these methods, Menaechmus detected that resolving the Delian problem of the duplication of the cube would require being able to outline three types of curves: the three types of conicals obtained with the same method, in other words, sectioning the three types of cone with a perpendicular plane to the generatrix, according to whether the angle at the vertex was acute, straight or obtuse. The three types of curves that Menaechmus had to draw were two parabolae and a hyperbola.



The three Menaechmus conics as a solution to the Delian problem of duplicating the cube. The length of the side of a cube with double the volume of another with a dimension of 7 metres should be 8.82 metres ($2 \cdot 7^3 = 8.82^3$).

The interjection point of the three curves can be found, with modern algebraic methods, resolving the system of three equations:

$$\begin{cases} x^2 = 7y \\ y^2 = 2 \cdot 7 \cdot x \\ xy = 2 \cdot 7^2 \end{cases}$$

that results in: $x^3 = 2 \cdot 7^3$.

Menaechmus obtained the value of x graphically, using the point of intersection of the three curves drawn, resulting in $x = 8.82$. If this Delian problem had been raised nowadays, with the algebraic tools we have at our disposal, the following equation would be used to solve it:

$$L^3 = 2 \cdot l^3.$$

With our current mode of writing mathematical calculations, the solution is:

$$L = \sqrt[3]{2} \cdot l.$$

For $l = 7$ m, using a scientific calculator, the result would be:

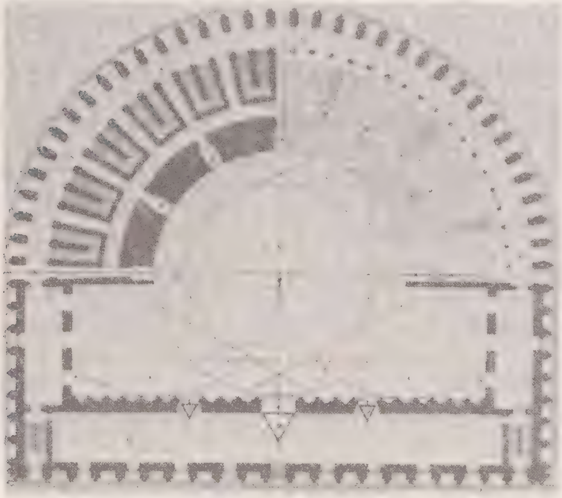
$$L = \sqrt[3]{2} \cdot 7 = 8.819447349.$$

We deduced that Menaechmus' approximation for the duplication of the Delos altar was quite good: $L = 8.82$. In decades after Menaechmus his conical curves were given rather prosaic names: 'section perpendicular to a generatrix of an acute-angled cone', or alternatively 'rectangled' or 'obtuse-angled'. By about 150 years after the work of Menaechmus the names ellipse, parabola and hyperbola were being used instead.

The first book of *Conics* was sent by Apollonius (c. 262–c. 190 BC) to Eudemus at Pergamon, a city some 25 km from the Aegean Sea, where today the Turkish town of Bergama stands today. Pergamon had been an independent state from the year 263 BC. It was a difficult time, with political treaties with a powerful Rome and battles with neighbouring states. Apollonius worked in Alexandria, the city with the largest library in the ancient world and the most advanced centre of research. In the dedication of his first book he comments on his life in Alexandria and that he had visited Pergamon. This illustrates the way of life of mathematicians of that era: they travelled, visited each other to exchange questions and ideas. In other dedications in volumes 2 and 3 to his friend Eudemus of Pergamon, he talks of other mathematical meetings in Ephesus. The work of Apollonius was sponsored by King Attalus I of Pergamon, to whom the five following volumes of the *Conics* were dedicated. In those, Apollonius demonstrated that it is possible to obtain three types of sections from a single cone, by varying the incline of the plane that cuts the cone. This represented an important step in the unification process of the study of the four curves because it avoided the cone having to be straight (with a vertex angle of 90°). Apollonius also used the two-sheeted cone for the first time and thus was able to draw the two arms of the hyperbola. Following a suggestion by Archimedes, the names *ellipse*, *parabola* and *hyperbola* were given to the conic sections – the circle is a special kind of ellipse. Although the concepts in mathematics have always been more important than the terms used, in this case the change of nomenclature had an extraordinary repercussion. The terms used by Apollonius came from the ancient Pythagorean language for solving second-degree equations: ellipse meant 'deficiency'; hyperbola, 'excess' (as in hyperbole, meaning exaggeration), and parabola comes from 'equalisation'.

Circles

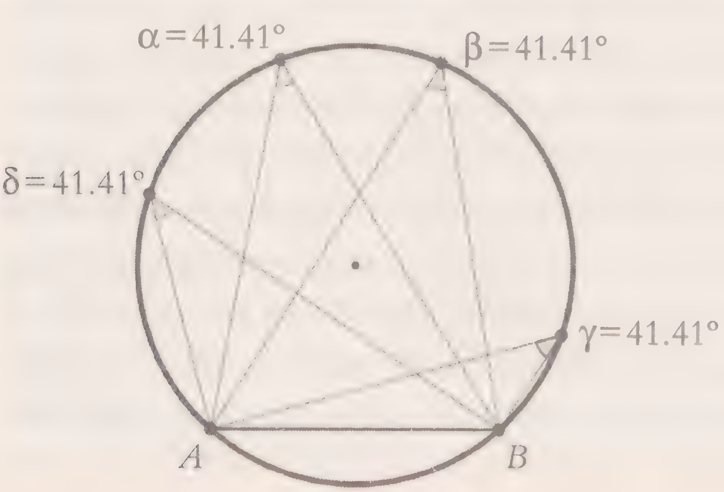
The circle is the curve defined as the set of all the points found at the same distance (the radius) from another point called the centre. It was the curve of the gods, the most perfect of curves. In the ancient world, extraordinary philosophers, such as Aristotle, attributed to it the form of the most sacred things, such as the orbits of stars, the form of the Earth and the Universe.



A Greek theatre; the circle has been the basic line of design and of the construction of theatres since ancient times.

Wheel of an Indian temple (10th century). The wheel was a invention of eastern civilisations around 3,500 BC, and it is based on the properties of the circle.

This curve is obtained as a conical section using a perpendicular plane to the cone axis. The eccentricity of the circle is 0. It can be considered to be an extreme of the ellipse when the focal distance is 0 and the axes a and b have the same length. It is the only curve where the tangent line at any point forms a right angle with the corresponding radius.



The circle is the only curve in which a chord (in this case labelled AB) projects the same angle (in this case, 41.41°) at any point on the circumference of the circle.

4,500 years ago the scribe Ahmose, the author of the Egyptian Rhind Papyrus, gave the length of the circle the value of

$$\frac{256}{81} = 3.16 \text{ times its diameter.}$$

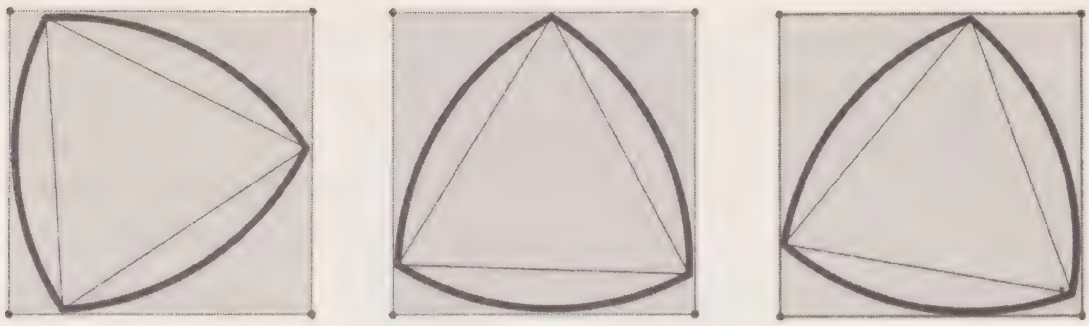
The Greeks Thales, Anaxagoras, Euclid and Archimedes continued the study of the circle based on Egyptian geometry and tried to resolve the issue of the inscription of different regular polygons in a circle and the problem known as squaring the circle – in other words, the geometric construction of a square that has the same area as a given circle. But it was Eudoxus, around the year 350 BC, who developed the depletion method or 'approaching the limit' and obtained a formula to calculate the area of the circle and the length of the circumference:

$$\text{Area} = \pi \cdot \text{radius}^2.$$
$$\text{Length} = 2\pi \cdot \text{radius}.$$

Curves of constant width

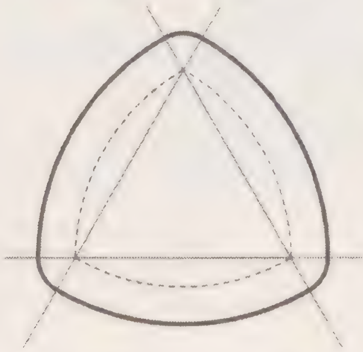
If we have to move a very heavy object from one place to another, perhaps it wouldn't be very practical to use wheels as they can warp or snap. The best solution appears to be using a flat platform supported on various cylindrical rollers: when moving forwards the rollers behind are picked up and placed at the front. This way, the load moves forward without being jolted. This is due to the fact that the roller has a section that is a circle and this curve is closed and at a constant width. Elliptic rollers could not be used, as the platform would lurch.

As the circle has the same width in all directions, it can rotate between two parallel lines without changing the distance between them. This might seem to be an exclusive property of the circle, but that is not the case. An infinity of closed curves exist with a constant width which are not circles. The simplest is the Reuleaux triangle, named after its inventor, the German mathematician and engineer Franz Reuleaux (1829-1905). It was built from an equilateral triangle (one with three equal sides and three equal angles) with a circular arc connecting each neighbouring two vertices. A vehicle rolling on wheels with this cross-section would provide as smooth a ride as any with cylindrical wheels. Drill bits have been built using the Reuleaux triangle that create square holes, such as the one designed by the American engineer Watts.

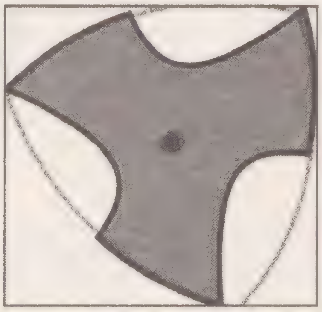


It is observed that in a fixed square the Reuleaux wheel can turn perfectly.

Reuleaux triangular wheel.



Curve with constant width and rounded corners.



Watts' drilling section for square holes.

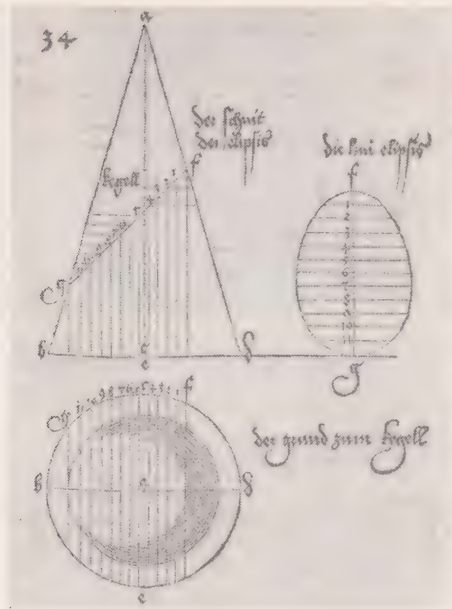
Ellipses

A complete circle viewed in perspective (or rather, in a direction that is not perpendicular to the plane on which it is found) is seen as an ellipse. In an ellipse, the sum of the distances from any point to any focal point is a constant quantity that is equivalent to the length of the greatest axis $2a$. The defining parameter of the more or less circular form of one ellipse is called eccentricity (e) and its value is:

$$e = \frac{c}{a} = \sqrt{1 - \frac{b^2}{a^2}} < 1.$$

The numbers a and b are half of the length of the largest axis and the smallest axis of the ellipse, while c is half the distance between focal points. Eccentricity of any circle is always 0 given that $c = 0$. The eccentricity of any ellipse is lower than 1 given that a is always greater than c.

The German painter Albrecht Dürer (1471-1528) gave his ellipses an ovoid form, as he thought that the cuts by the oblique plane to a cone should produce lower curvature in the upper part of the section. This error made Kepler laugh when contemplating it in Dürer's book, *Underweysung*, published in 1525.

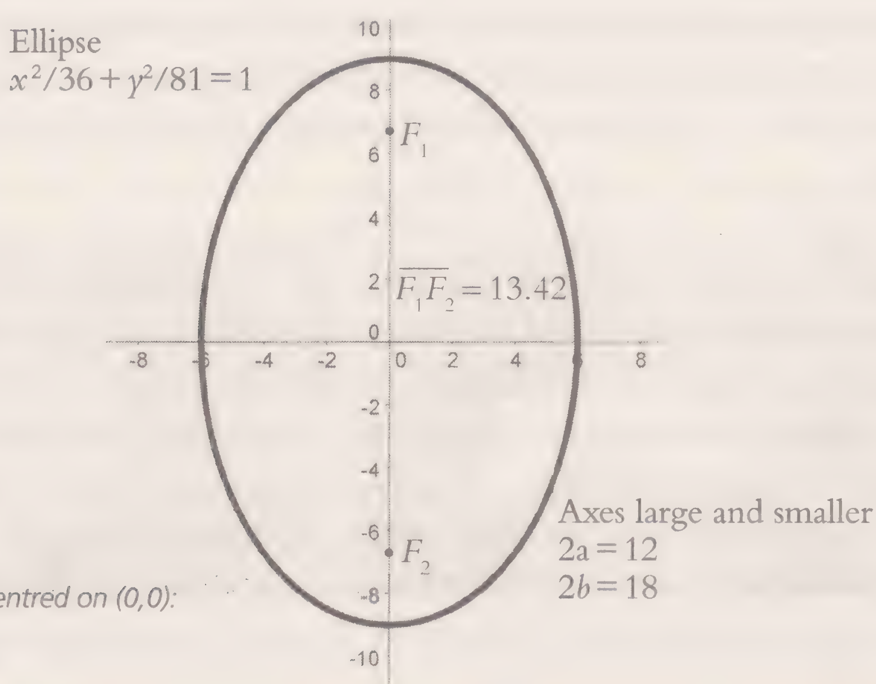


The method using a ruler and compass used by the German painter Albrecht Dürer to draw ellipses. This was based on Conics *by Apollonius.*

The focal points of the ellipse in the following figure are indicated as: $F_1\ (0,c)$ and $F_2\ (0,-c)$. Throughout the ellipse, a relationship is fulfilled between a, b and c: $a^2 = b^2 + c^2$. The value $2c$ is called the focal distance. Therefore, in this ellipse,

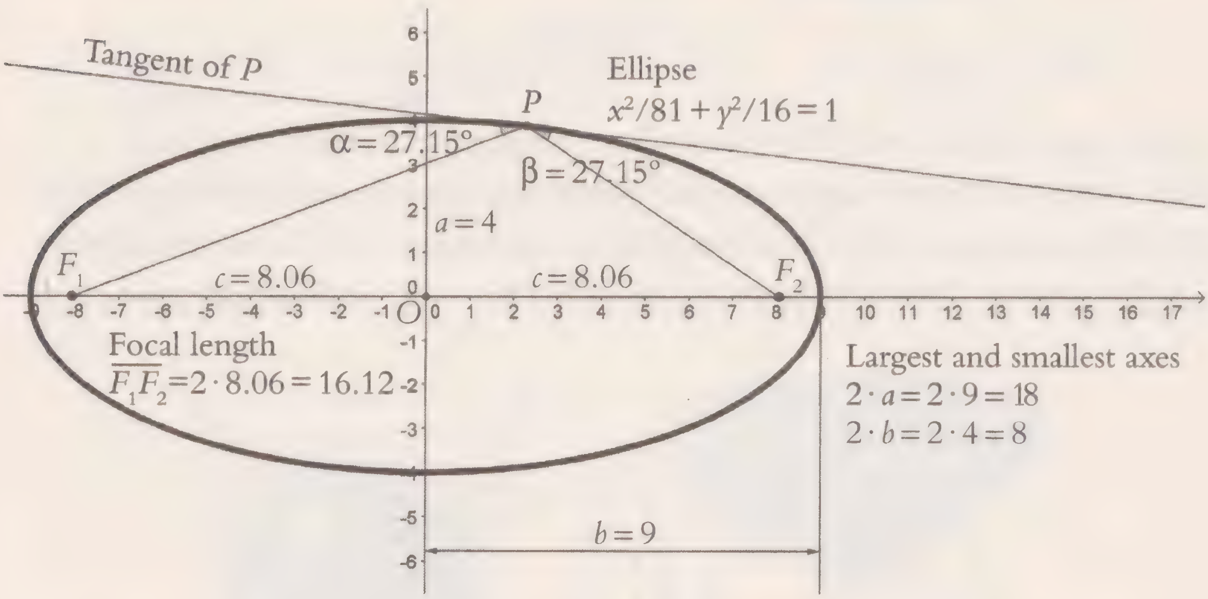
$$9^2 = 6^2 + c^2; \quad c^2 = 9^2 - 6^2 = 81 - 36 = 45; \quad c = \sqrt{45} = 6.71,$$

and the focal distance is: $\overline{F_1 F_2} = 2c = 13.42$.



An ellipse centred on (0,0): $\dfrac{x^2}{a^2} + \dfrac{y^2}{b^2} = 1.$

It is said that Lewis Carroll (1832-1898), pastor, mathematician and writer of *Alice in Wonderland*, constructed a circular billiards table. If instead of being circular it had been elliptical, when a ball went through a focal point it would always bounce back through focal point, and this would be repeated until the ball stopped. A ball hit on an elliptic billiards table would bounce off the edge of the curve as if it had hit a perfect wall that was in the place of a tangent to the curve at that point. Due to this property, a ball launched from a focal point F_1, 'bounces off' the tangent line leaving equal angles α and β on each side (like any reflection) and then moves towards another focal point F_2, as is shown in the diagram.



This property also applies to light. Apollonius was the first to demonstrate that when placing a source of light at the focus of an elliptic mirror, the light reflected in the mirror would concentrate on another focal point.

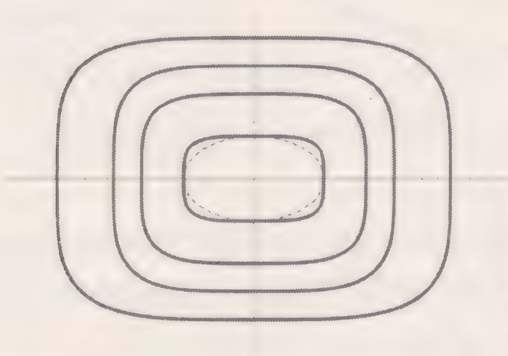
The ellipse appears in contemporary architectural constructions, such as the Auditorio *by Calatrava in Tenerife (left), or in public buildings such as those by Niemeyer in Brasilia.*

Super ellipses

Some curves exist that are positioned between the ellipse and the rectangle; they are known as 'super ellipses'. They were created in 1945 to rectify a problem that had arisen in the urbanisation of a square in Stockholm. Two large avenues were to cross forming an enormous rectangle. When planning it, the Swedish architects couldn't use the ellipse as a form for the square, as its extremes were too acute and would hinder the road traffic around it. To solve the problem the Danish writer and inventor Piet Hein (1905-1996) was consulted, who presented an ingenious mathematical solution. The equation of an ellipse is:

$$\frac{x^2}{a^2}+\frac{y^2}{b^2}=1;$$

What type of curve would be obtained on increasing the exponents to 2, 3, 4,....? The result was a much more rounded ellipse, which looked more like a rectangle. When the exponent is increased indefinitely (tending to ∞), the super ellipse becomes converted into a rectangle of sides *a* and *b*.



The equations of the four super ellipses are:

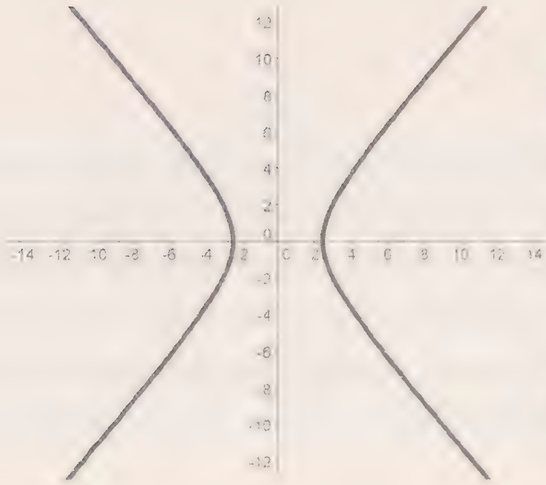
$$\frac{x^4}{5^4}+\frac{y^4}{3^4}=1;\quad \frac{x^4}{8^4}+\frac{y^4}{6^4}=1;\quad \frac{x^4}{10^4}+\frac{y^4}{8^4}=1;\quad \frac{x^4}{14^4}+\frac{y^4}{10^4}=1.$$

The initial ellipse (outlined) is: $\frac{x^2}{5^2}+\frac{y^2}{3^2}=1.$

As in the anecdote of 'Columbus' egg', Piet Hein discovered that a solid body with a 'super ellipsoidal' form comes to rest in a vertical position from any initial position.

Hyperbolae

The hyperbola is a conical curve with a centre that has two separate arms. Its eccentricity is greater than 1. The trajectories of some celestial bodies attracted by a centre of gravitational attraction can be hyperbolic, although they are among the most rare trajectories.

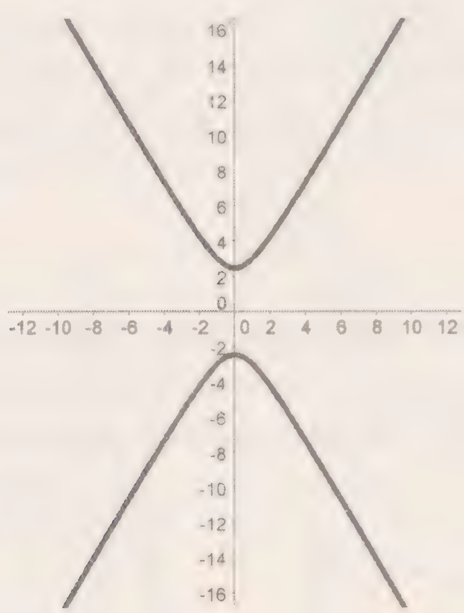


A hyperbola that when rotating around axis Y produces the 3D hyperboloid of one sheet.

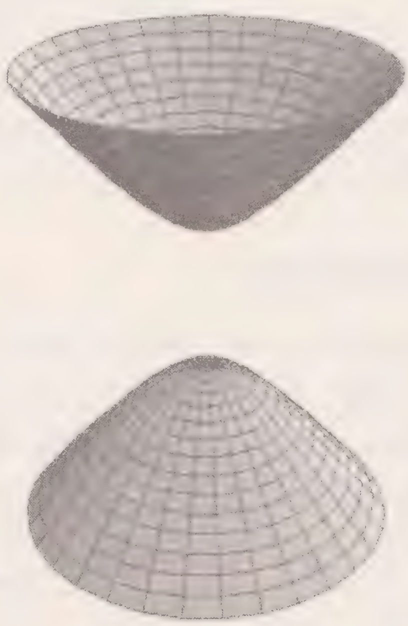
One-sheeted hyperboloid:

$$\frac{x^2}{a^2}+\frac{y^2}{b^2}-\frac{z^2}{c^2}=1.$$

The great Greek mathematician Euclid wrote about the hyperbola but he only considered one of the two arms, although the first one to study it was Menaechmus, as we have already seen.



The two arms of a hyperbola rotating around axis Y produce a 3D body called the hyperboloid of two sheets.



Two-sheeted hyperboloid:

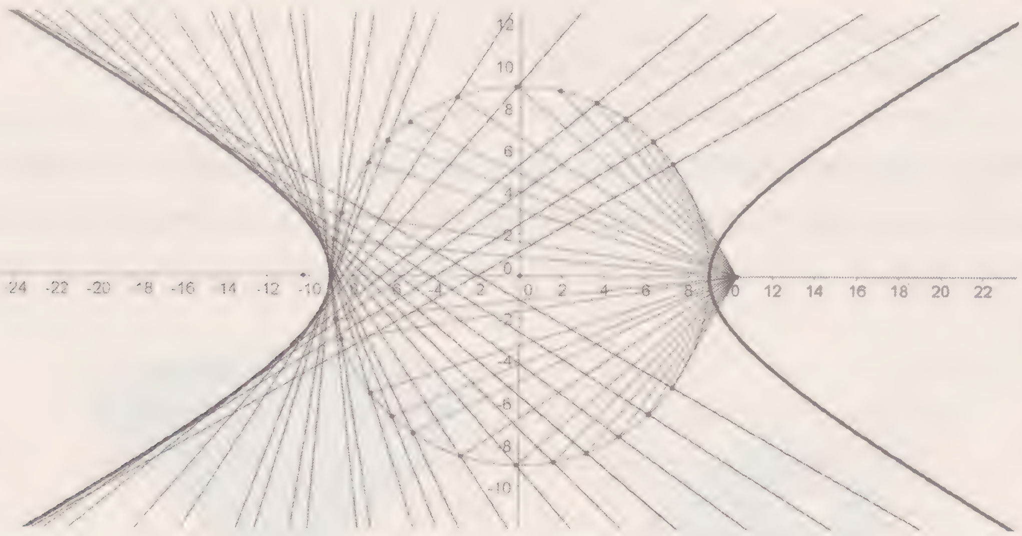
$$\frac{x^2}{a^2}+\frac{y^2}{b^2}-\frac{z^2}{c^2}=-1$$

The line that passes through the focal points of the hyperbola cuts the same at the points called vertices. The distance between the two vertices of the following hyperbola:

$$\frac{x^2}{a^2}-\frac{y^2}{b^2}=1$$

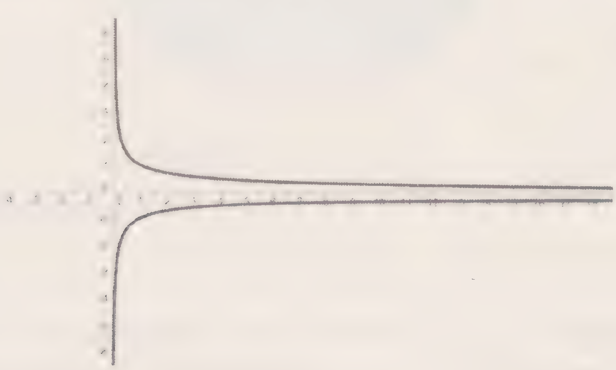
is the greater axis of the same, which measures $2 \cdot a$. The centre of the hyperbola is the mid-point of this segment; in this case, it is the origin of the coordinates (0,0).

A set of lines is drawn from a focal point that cuts the circle of radius a and centres on (0,0) at a series of points. At the cutting point of each line another line is drawn perpendicular to it. Many lines are obtained that determine a curve that is said to be enveloping the family of lines. This curve forms one of the hyperbola arms, as can be seen in the figure:

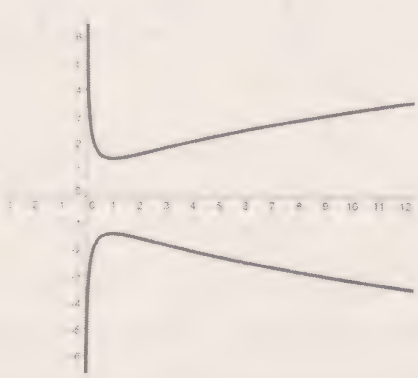


Cubic hyperbolae

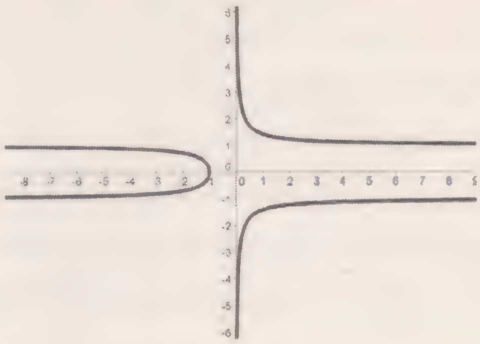
A curious family of curves exists called cubic hyperbolae, which have a general equation $x \cdot y^2 = P(x)$, where $P(x)$ is a polynomial of a degree equal to or lower than 3. They were studied by Isaac Newton in 1701, who gave them the name of 'semi-tridents' due to their form.



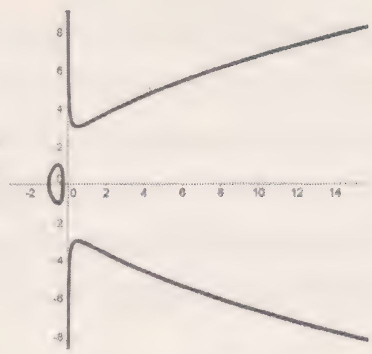
$x \cdot y^2 = 1.$



$x \cdot y^2 = 1 + x^2.$



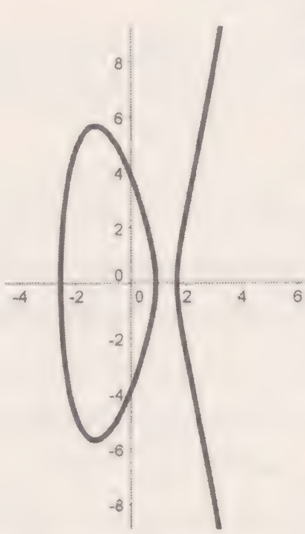
$x \cdot y^2 = 1 + x$.



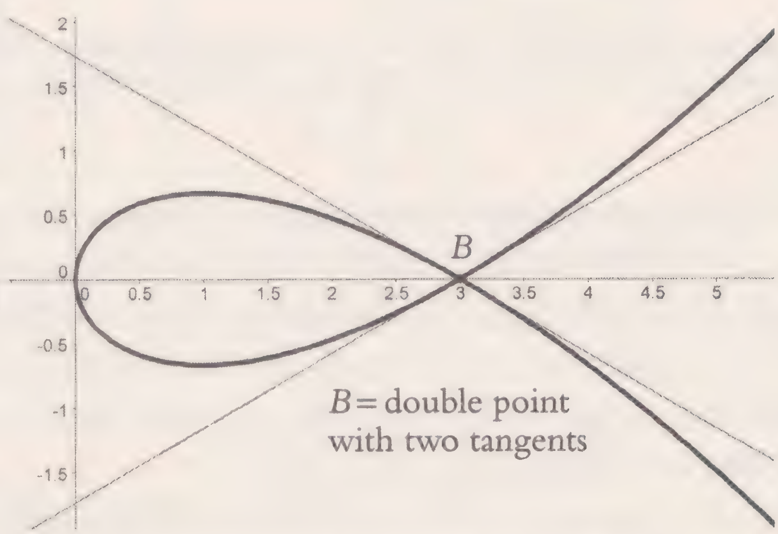
$x \cdot y^2 = (x+1)(2x+1)$.

Curves $x \cdot y^2 = 1$ and $x \cdot y^2 = 1 + x^2$ are cubic, or rather, of degree 3. All have a singular point, a vertical asymptote at point (0,0); there the curve 'soars towards infinity'. In the first, the curve reduces towards zero when the values of x become very large. In the second, in contrast, the curve opens gradually and the function increases more and more towards positive and negative infinite numbers.

Curves $x \cdot y^2 = 1 + x$ and $x \cdot y^2 = (x+1)(2x+1)$ also have a singular asymptote point at (0,0). The first has a zero in the negative value of x ($x = -1$) and two horizontal asymptotes (x jump from $+\infty$ to $-\infty$) at $y = +1$ and $y = -1$. The second has two zeros in the oval. There are many other cubic curves, but there are two that are very special – the elliptic cubic and Tschirnhausen's cubic curve, represented as follows:



Elliptic cubic curve:
$y^2 = 4x^3 - 20x + 15$.



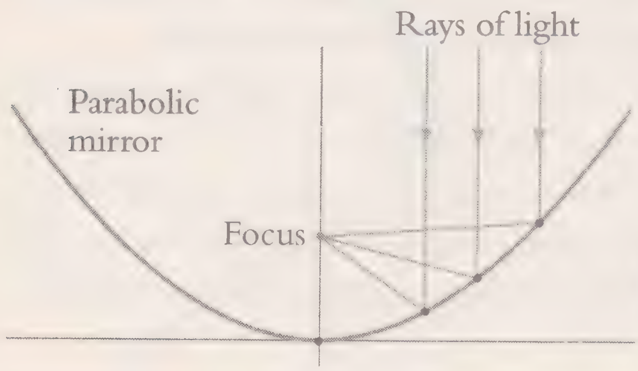
Tschirnhausen's cubic curve: $9y^2 = x(x-3)^2$.

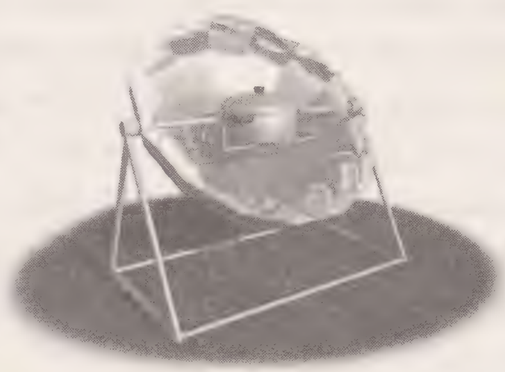
The elliptic cubic curve has three real zeros, or in other words, it cuts the X axis three times, twice with the oval and once with the arm.

The German mathematician and doctor Ehrenfried Tschirnhausen (1651-1708), who invented the formula for making porcelain in Europe (it had previously been developed in China, but the ingredients were not known), proposed the cubic that carries his name in his mathematical works. This curve has a zero at point (0,0) and another, which is double, at point B (3,0); it is double because it cuts twice at (3,0) through axis X, one cut by each arm. Furthermore, point (3,0) is special because it has two different tangents to the curve depending on the arm being considered.

Parabolae

The parabola has some very interesting reflective properties. If a parabolic mirror receives light from a distant source, which means that the incident beams are parallel to the mirror axis, then the reflected light concentrates on the focal point.

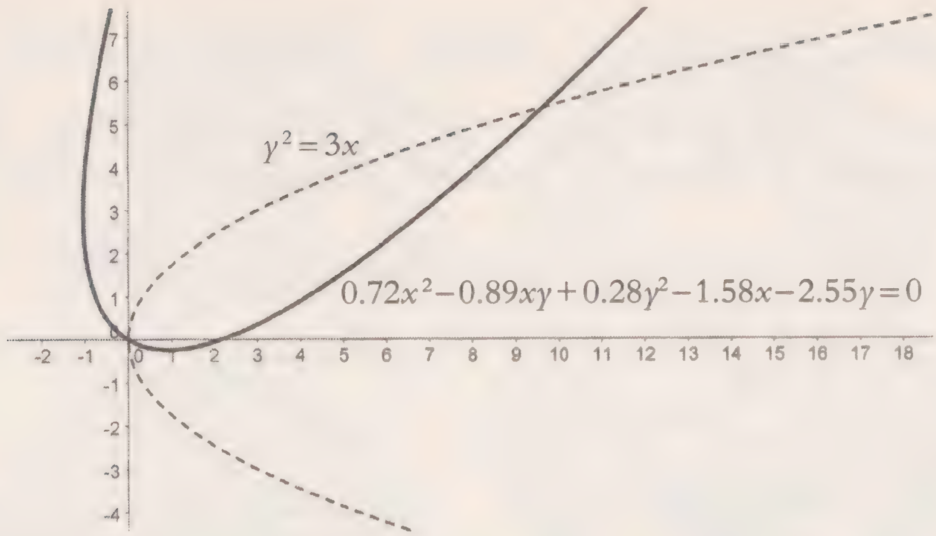




The diagram on the left demonstrates how the beams that are parallel to the mirror axis focus on the same point. This property is used in many fields of everyday life, such as in solar ovens (right).

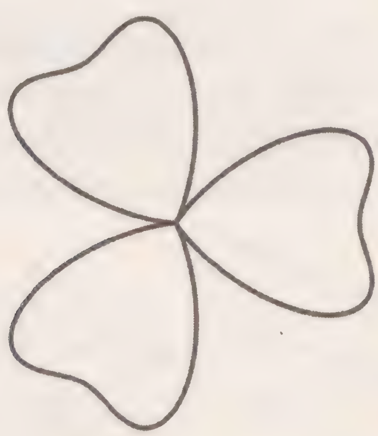
Legend says that Archimedes prepared for the Roman siege of the city of Syracuse by making use of this parabolic reflection property. He constructed a system of metallic mirrors placed in a parabola, which focused the Sun's beams on the Roman fleet, setting fire to their ships. This property of the parabola has a multitude of modern uses in radar, television antennae, automobile headlights and solar ovens, among others.

When a parabolic curve that passes through the origin (0,0) has a symmetrical horizontal axis, its equation is like the dotted parabola in the figure $y^2=3x$. When it has a symmetrical oblique axis, its equation is more complicated, like the parabola with the axis at 45° in the following figure, which has the equation (in implicit form): $0.72x^2-0.89xy+0.28y^2-1.58x-2.55y=0$.

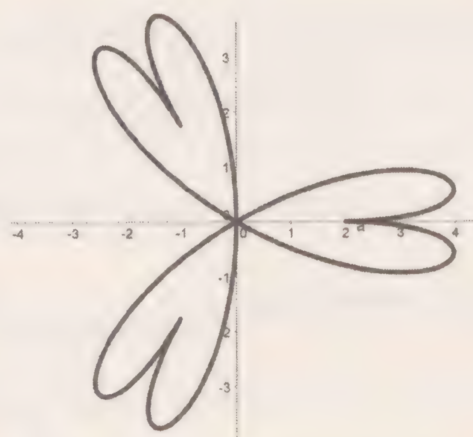


Attractive curves

In this section a series of curves are included that are used as decorative motives, and at the same time they are forms that are surprisingly similar to objects existing in nature. And their equations are known!



Habenicht clover. Equation in polar coordinates: $r=1+\cos 3\theta+\text{sen}^2 3\theta$.

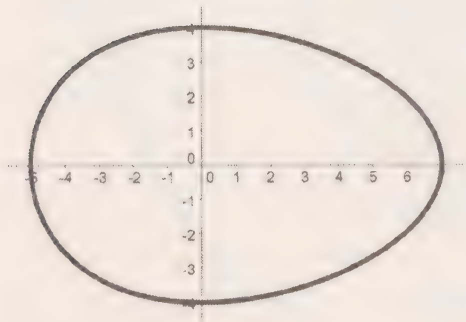


Brocard clover (algebraic curve).
Implicit equation in Cartesian coordinates:

$$x^2+y^2-\frac{4x(x^2-3y^2)}{x^2+y^2}+4-\left(x^2+y^2-\frac{3x(x^2-3y^2)}{x^2+y^2}+2\right)^2=0.$$

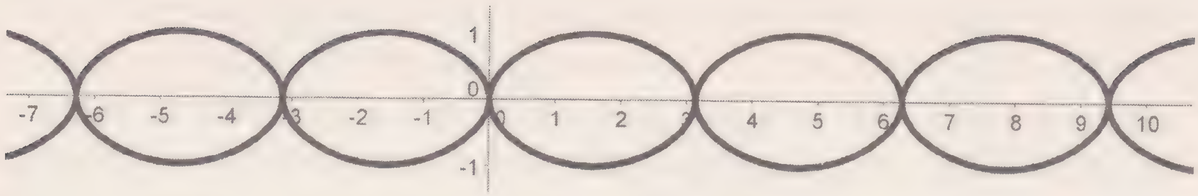
Equation in polars:

$$r=\frac{|\sin 3\theta|\cdot((\cos 3\theta+1)^{1/3}-(1-\cos 3\theta)^{1/3})}{(1-\cos 3\theta)^{1/6}\cdot(\cos 3\theta+1)^{1/6}}+2\cdot\cos 3\theta.$$

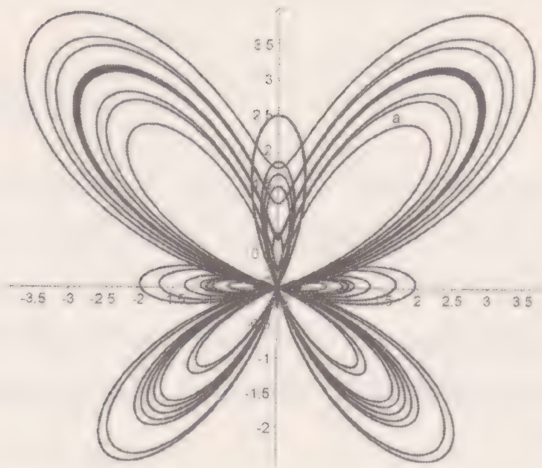


Hügelschäffer egg. Equation in Cartesian coordinates:

$$f(x)=\pm 4\sqrt{\frac{36-(x-1)^2}{35+2x}}\ .$$

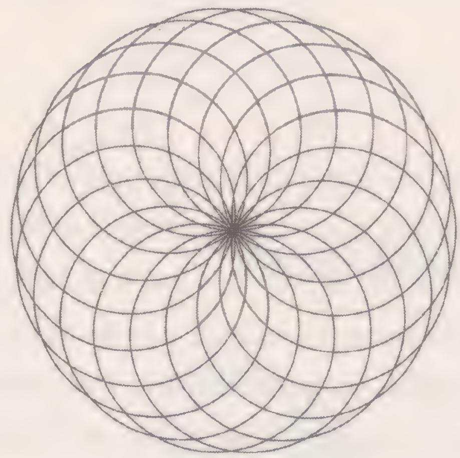


Chain of eggs; the equation in Cartesian coordinates is $f(x) = \pm\sqrt{|\sin(x)|}$.



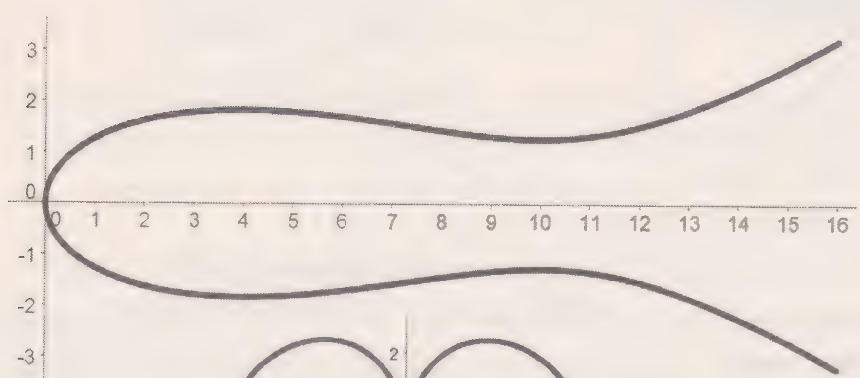
T. Fay butterfly; equation in polar coordinates:

$$r = e^{\cos\theta} - 2\cos(4\theta) + \sin^5\left(\frac{\theta}{12}\right).$$



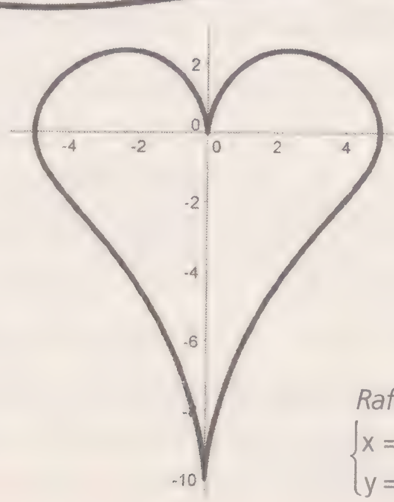
Moritz curve; equation in polar coordinates:

$$r = \cos\frac{9\theta}{10}.$$



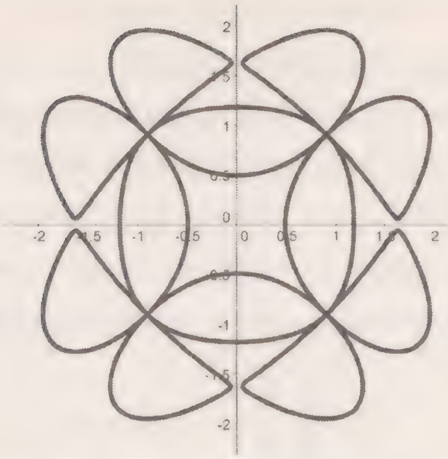
Fish curve; equation in Cartesian coordinates:

$$y = \pm\sqrt{\frac{2x^3 - 42x^2 + 240x}{125}}.$$



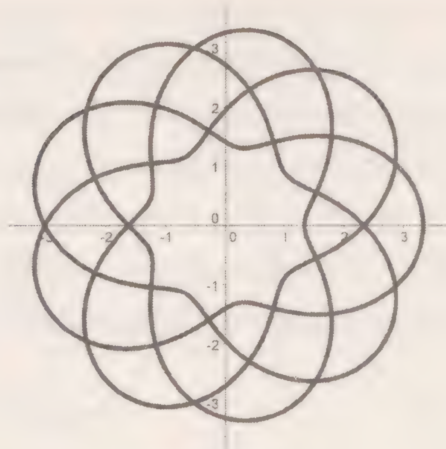
Rafael Laporte heart:

$$\begin{cases} x = 5\sin^3(t) \\ y = 5\cos(t) - 5\cos^4(t) \end{cases}$$



Symmetrical curve of the rosace with four arms; the parametric equations:

$$\begin{cases} x = \pm\sqrt{3\cos 2t \cdot \cos t + 0.82} \\ y = \pm\sqrt{3\cos 2t \cdot \sin t + 0.82} \end{cases}$$

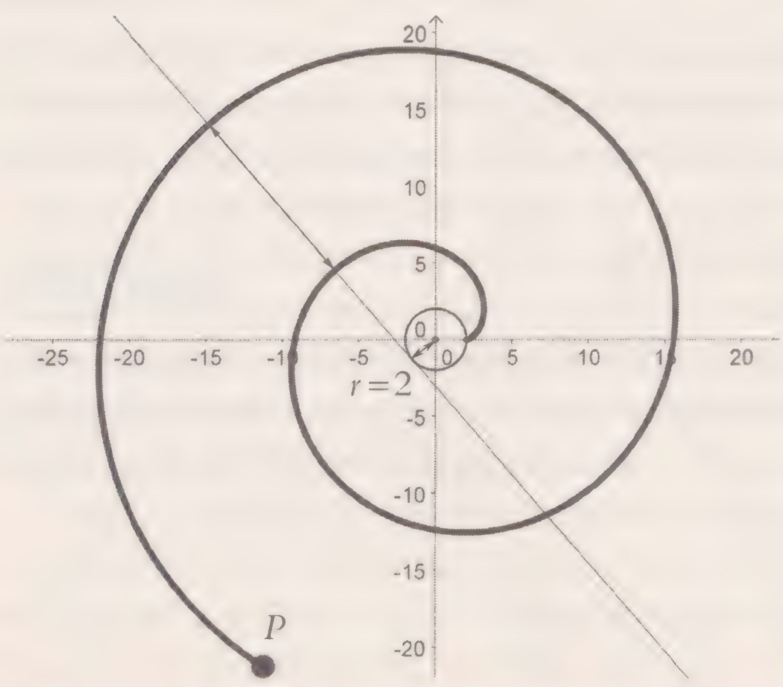


Ornamental curve; equation in polars:

$$r = \cos\frac{9\theta}{4} + \frac{7}{3}.$$

Involute of a circle

This spiral curve is obtained by tracing the trajectory of one end of a taut thread as it is wound around a circular spool held in a fixed position. The fixed point of the centre of the spool is (0,0), the spool has a radius of 2 cm and the distance between each iteration of the involutes is constant and measures $4\pi \cong 12.57$cm.



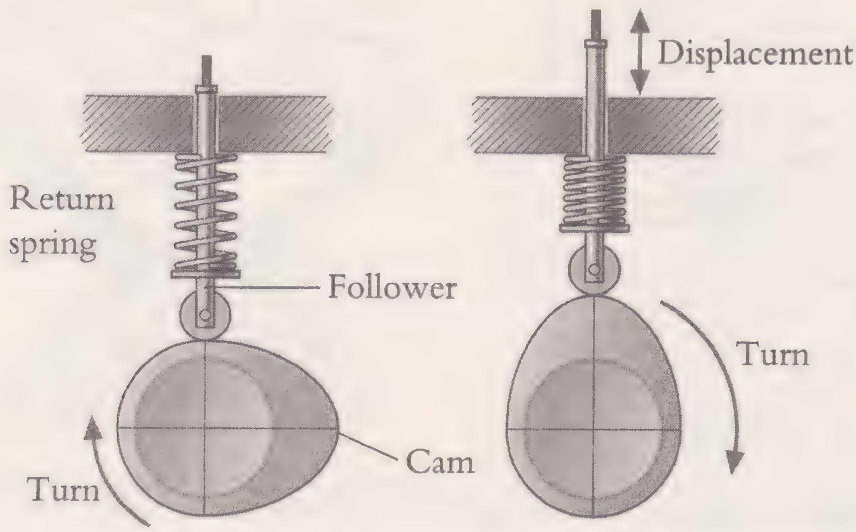
Parametric equations of the circle involute:

$$\begin{cases} x = 2(\cos(t) + t \cdot \sin(t)) \\ y = 2(\sin(t) - t \cdot \cos(t)) \end{cases}$$

Polar equation:

$$\theta = \pm\frac{1}{2}\sqrt{r^2 - 4} - \operatorname{acos}\left(\frac{2}{r}\right).$$

When it has been turned several times (t is now elevated), the involute tends to be confused 'asymptotically' with the Archimedean spiral, and for that reason it is often used in technology as the most suitable profile for the cams of an engine. The geometric property that explains the use of the curve as a profile for designing gears is clearer if we think of it as the opposite to the way the curve has been defined. If the involute is made to rotate around its centre, the tangents of the curve direct motion in both directions thanks to a return spring, as can be seen in the illustration:

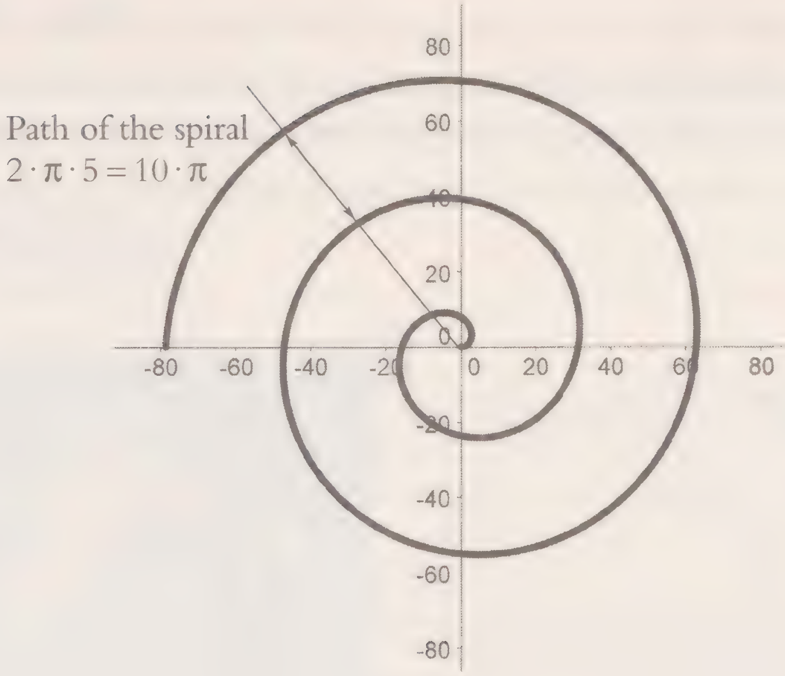


A cam mechanism formed with two arc of an involute of a circle.

Archimedean spiral

The Archimedean spiral was studied by the Greek mathematician in the year 225 BC in his work *On Spirals*, where he used it to solve two of the three ancient Delian problems – the trisection of the angle and the squaring of the circle. This spiral is the curve of the trajectory from a point that moves uniformly along a line of a plane while the said line rotates uniformly around one of its points; it is the groove of an old-fashioned vinyl music record. The point (0,0) is the centre of the rotation, and it is a case limit of the involute of the circle with radius $r = 0$ for the initial angle $\theta = 0$. This is a transcendent type of curve. The one in the following figure has an equation in polar coordinates of $r = 5\theta$ and a passage of $10\pi \cong 31.42$. Its equation in parametric coordinates is:

$$\begin{cases} x = 5 \cdot t \cdot \cos(t) \\ y = 5 \cdot t \cdot \sin(t) \end{cases}$$



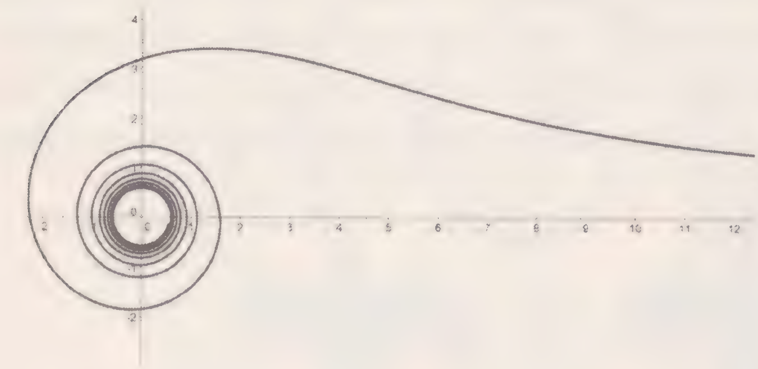
This spiral has some very special properties. For example, the area of the spiral in its first bend is equal to a third of the area of the circle that surrounds it. The fact that the spiral is a simple form to construct explains why it has been used for ornamental purposes since ancient times. To make one on a potter's wheel, all you need to do is to move the finger or brush in a certain direction from the centre towards the edge at a constant speed while the wheel turns. They have found spirals in Bronze-Age funerary burial mounds, in Greek and Etruscan pottery and in other ceramics as a decoration on plates or inscriptions. The Celtic culture developed ornaments and medallions in which they were represented, using three spirals that weave in and out of a circle turning in both directions – the duality of the forces that are in permanent interaction in nature and the equilibrium expressed by the number three.

Other spirals

The word *spiral* refers to a flat curve that is outlined with the movement of a point that winds around a fixed pole with continuous forward and backward movements. Spirals are always transcendental curves.

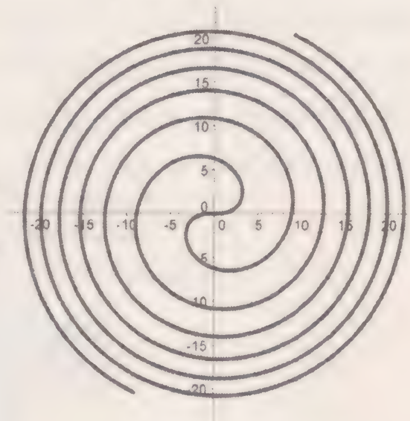
There are innumerable natural manifestations of spirals, whether organic, artistic or mechanical in nature. For that reason these curves cannot fail to draw the attention of mathematicians. However, as their own form suggests, they are curves that 'sidestep'. They are not geometrically static curves, such as the circle or parabola, but rather

to construct spiral curves certain mechanical resources are needed. In other words, they are objects that grow or that move. According to their 'mathematical' definition, they are flat curves that start at a point and whose curvature progressively decreases.

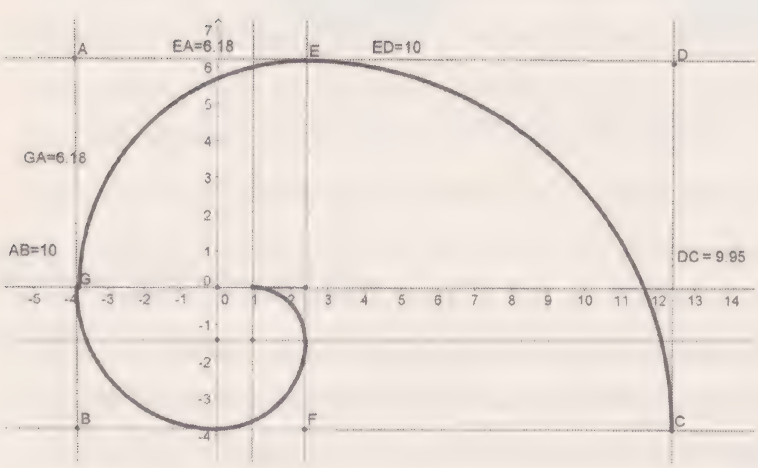


Lituus spiral or 'bishop's crozier' (Cotes and MacLaurin, 1722). Polar equation:

$$r=\sqrt{\frac{16}{\theta}}.$$



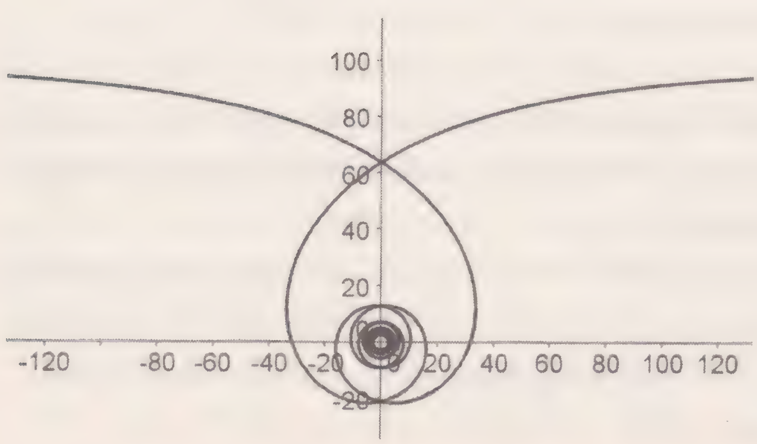
Fermat's spiral (1636). Equation: $r=\pm 5\sqrt{\theta}$. Curve with central symmetry with respect to the pole (0,0). Its inverse curve is the lituus.



Dürer's spiral (1525), based on the golden ratio:

$$\Phi=\frac{1+\sqrt{5}}{2}=1.618.$$

There are various tangents to the circular arcs.

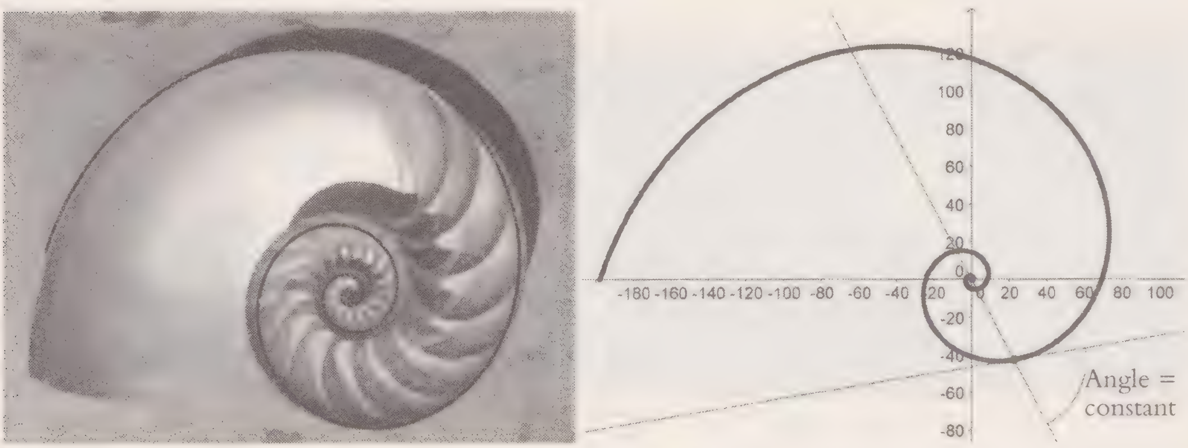


Hyperbolic spiral or Varignon's spiral (1704). Polar equation:

$$r=\frac{100}{\theta}.$$

Logarithmic spiral

This curve was studied first by Descartes and Torricelli in 1638, and later by Jakob Bernoulli (1654-1705). It is also known as an equiangular spiral and Bernoulli's spiral, who called it *spira mirabilis*. It appears repeatedly in nature in the forms of various animals and plants.



Left, a nautilus, a mollusc with a shell that approximates to a logarithmic spiral, represented on the right, with the polar equation being $r = -3 \cdot e^{\frac{\theta}{3}}$.

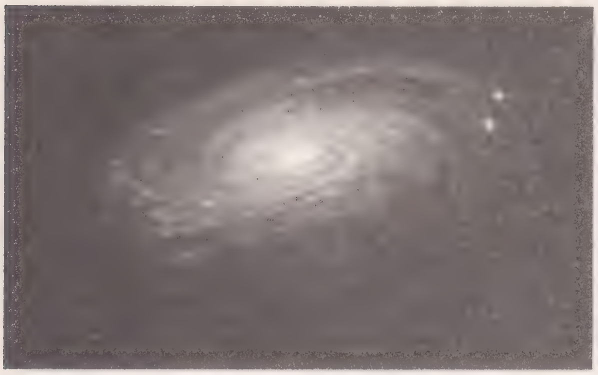
In general, the equation of any logarithmic spiral is $r = a \cdot e^{\frac{\theta}{k}}$. The one represented in the above figure has the values $a = -3$ and $k = 3$. It can be defined as the curve whose tangent-radius angle is constant, and this angle is the one whose tangent is the constant k. It is also the flat development of a spiral of a revolving cone.

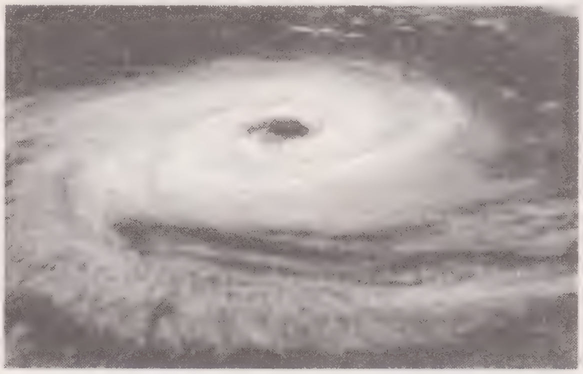
The logarithmic spiral presents exceptional 'stability' with respect to itself in the classic geometric transformations. This characteristic property implies that on rotating, extending, reducing, inverting or on drawing its evolute, its section, it always results in another logarithmic spiral.

This curious property impacted Bernoulli considerably and it inspired the engraving that he had made on his gravestone in the cemetery of Basel, his dear *spira mirabilis* with the inscription *Eadem mutata resurgo* ("Although changed, I shall arise the same").

In the so-called 'pursuit curve" problem, four animals that run at a constant speed start at the four vertices of a square. Each one aims to reach the animal positioned to its right. The curves followed by the four animals are logarithmic spirals that are found in the centre of the square.

The changes in orbit of a spaceship are made following a transference orbit in the form of a logarithmic spiral that joins two circular orbits of a different radius. In order for the craft to follow this trajectory it only requires a engine that provides a small acceleration throughout the journey and that reduces as the spaceship moves away from the centre of the Earth. Spiral curves also appear in the forms of galaxies and hurricanes, as can be seen in these photographs:





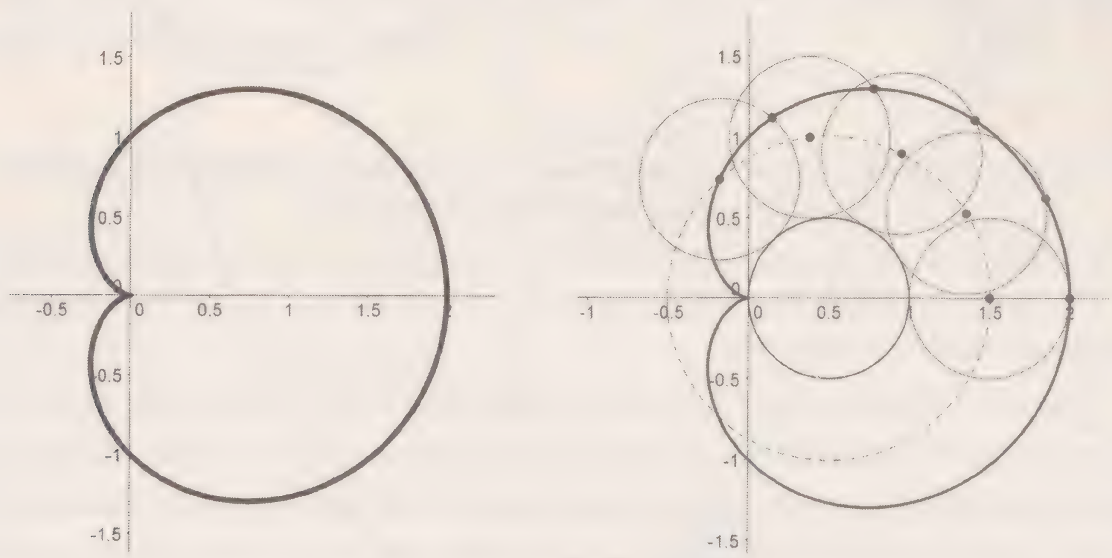
Photograph of the spiral galaxy Messier 88 (left) and Hurricane Katrina seen from the International Space Station (source: NASA).

The application of the logarithmic spiral is very important in classical and contemporary architecture. In modern sculpture and smithery, for example, many spirals can be found.

Double spiral staircase, Vatican Museums (Bramante, 16th century).

Cardioid

The name cardioid comes from the heart shape of this curve, which belongs to the family of curves called 'Pascal's snail' or *limaçon* in honour of its discoverer, Étienne Pascal (1588-1651), father of Blaise Pascal (one of the first to calculate probabilities in 1654 and designer of the first mechanical calculator). E. Pascal divulged his discovery in the wealth of correspondence he maintained with other European mathematicians and at the mathematical salon held by Marin Mersenne in his Parisian monastery – a social circle that had a great influence on scientific development at that time.

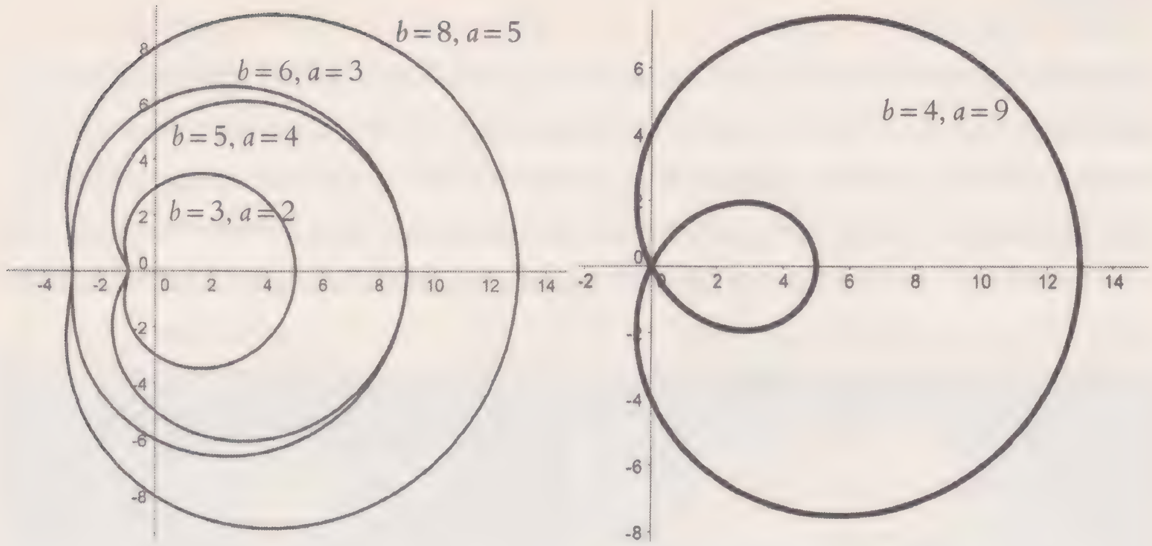


The Cartesian equation of the represented cardioid is $(x^2+y^2-x)^2=(x^2+y^2)$; its polar equation, $r=1+\cos\theta$, and its parametric equations:

$$x=(1+\cos(t))\cdot\cos(t)$$
$$y=(1+\cos(t))\cdot\sin(t).$$

This cardioid has a special type of special point at (0,0), mathematically called the 'regression point'. A cardioid like $r=2a(1+\cos\theta)$ is generated by making rotations, without it slipping, around a circle with radius a around another equal (fixed) one with its centre at point $(a, 0)$. In the cardioid considered here it is $a=0.5$. Hence, the fixed circle has its centre at (0.5, 0) and radius of 0.5.

The position of the point of the mobile circle with radius 0.5 which rotates around the fixed one, without slipping, generates all the points of the cardioid. It is deduced, therefore, that the cardioid is a particular case of the epicycloid curve described in Chapter 3.



Two types of limaçons *for different values of* b *and* a. *The one on the right-hand side is obtained when* b<a, *and those on the left, when* b>a.

The general polar equation of a *limaçon* is $r=b+a\cos\theta$, and its Cartesian equation is, $(x^2+y^2-2ax)^2=b^2(x^2+y^2)$.

This curve had already been studied by Dürer in 1525, who initially christened it the 'arachnid'. It also appears in more amusing situations, such as the following: A girl, whose waist is a perfect circle, is playing with her hula hoop, the diameter of which is double her waist. A mark is made on the hula hoop with a felt-tip pen. If asked what type of curve the trajectory of the point makes while the girl rotates the hula hoop around her, the answer may surprise a few – it is a cardioid!

Catenary

The catenary or 'chain curve', is the geometric form that an inextensible thread adopts (it doesn't stretch under its own weight) when suspended from two points and hangs down only subject to its own weight. Galileo was not right when he considered that the solution to this problem took the form of a parabola. Around the year 1690, Leibniz, Huygens and Johann Bernoulli established a correct formulation for this curve. In the equation of the catenary curve, there is a special function called hyperbolic cosine and it is related to the exponential functions:

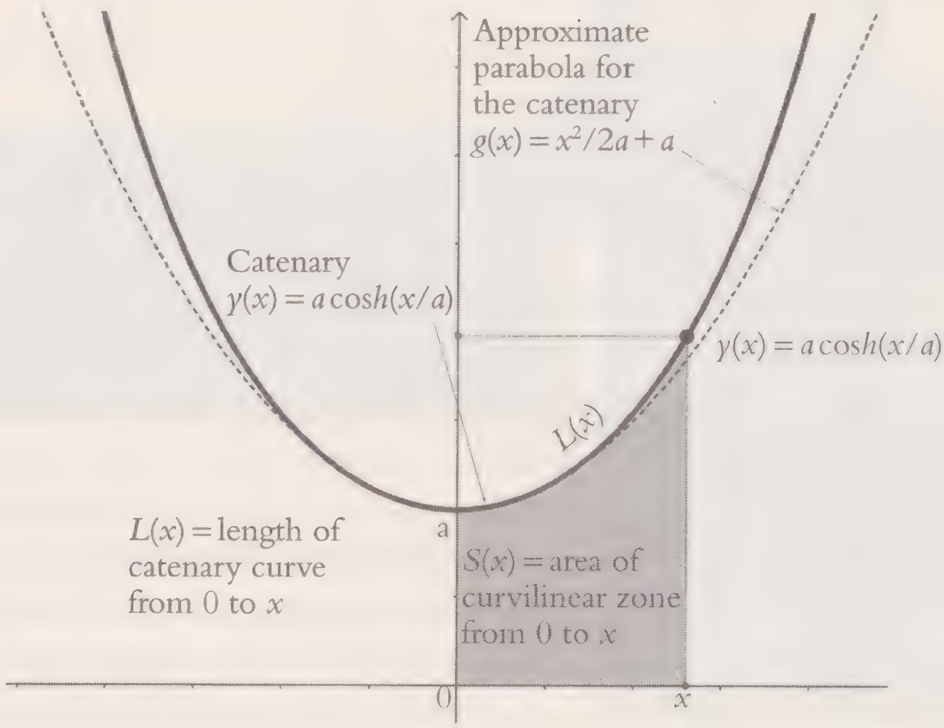
$$\cos h(x)=\frac{e^x+e^{-x}}{2}.$$

The corresponding function to the trigonometric sine is the hyperbolic sine which is defined:

$$\sin h(x) = \frac{e^{x} - e^{-x}}{2}.$$

Unlike the trigonometric sine and cosine of an angle x, which are related through the expression $\cos^2(x) + \sin^2(x) = 1$, the hyperbolic sine and cosine are related through the expression: $\cos h^2(x) - \sin h^2(x) = 1$.

In Chapter 4 the behaviour of the trigonometric and hyperbolic functions and curves were explained thoroughly.



Catenary: $y(x) = a \cdot \cosh\left(\frac{x}{a}\right)$. *Approximate 2nd degree parabola:* $y = \frac{1}{2a}x^2 + a$.

The equation of the catenary in parametric coordinates is:

$$\left.\begin{aligned} x &= a \cdot \ln(t) \\ y &= \frac{a}{2} \cdot \left(t + \frac{1}{t}\right) \end{aligned}\right\} \text{with } t > 0.$$

It can be deduced that the natural logarithm that appears in the equation comes from the inverse function of the basic exponential function number e. It should be

noted that $y=e^t$; $y=\ln(t)$ are inverse functions. The length of the catenary arc $L(x)$ measured from its lowest point ($x=0$, $y=a$) up to point (x,y) is:

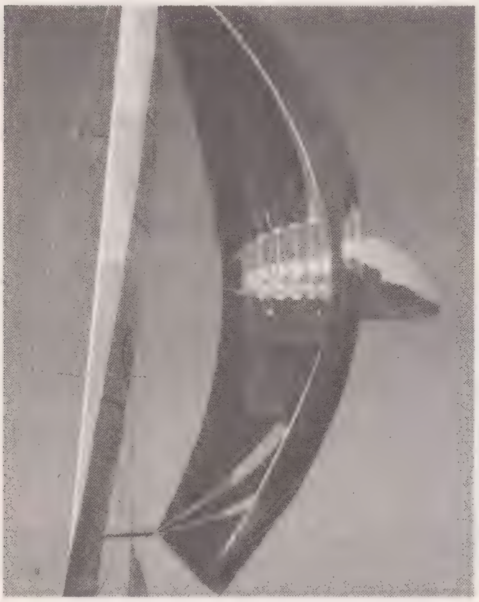
$$L(x)=a\cdot \sin h\left(\frac{x}{a}\right).$$

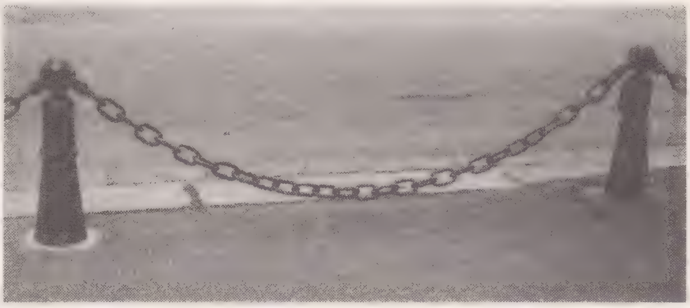
The relationship of the arc length $L(x)$ with its extreme ordinate $y(x)$, is:

$$y^2(x)-L^2(x)=a^2.$$

The area of the mixtilinear rectangle (with one curved side) formed by the curve, the abscissa axis, the ordinate axis and the vertical at the point of the considered curve, has the value:

$$S=a\cdot L(x).$$



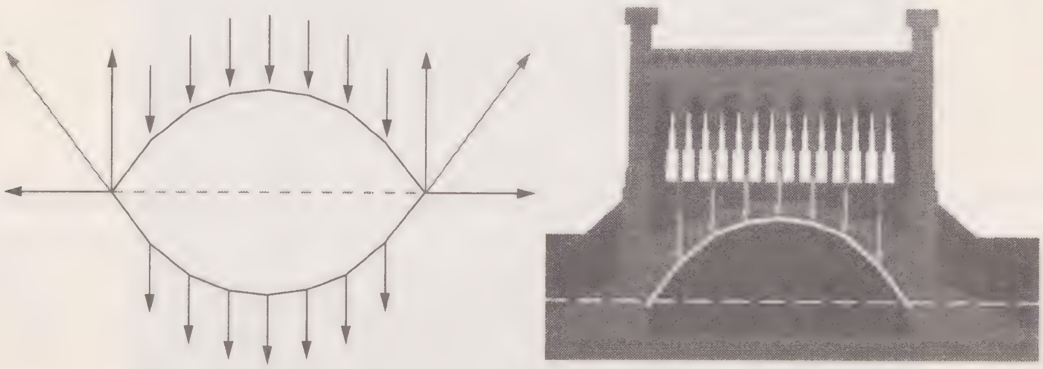


The transverse section of the sails of a yacht inflated by the wind is also a catenary, because the horizontal force that the wind exerts on the sail is similar to the action of the force of gravity on a chain, as can be seen in the photographs. This curve was called 'velarium' by Jakob Bernoulli.

If a parabola rotates along a straight line, the focal point of the parabola follows a catenary. Close to the vertex, the parabola and catenary are practically overlapping, as can be seen in the figure on the previous page, although the parabola will present a larger arrowhead than the catenary.

The catenary has the property of being the geometric location of the points at which the horizontal tensions of a cable are compensated. As such, the cable does not have lateral tensions and it does not move outwards. The tensions are divided between a vertical force (Earth's gravity) and a tangential tension to the cable at each point, which keeps it static.

Given a linear structural element in architecture subject only to vertical loads, the catenary form is the same as the curve that the gravity axis adopts (the axis of symmetry of the element's section) to minimise the tensions in the structural element. That is why this property is used for designing arches. An arc in the form of an inverted catenary is precisely the curved form that minimises compression forces in an arch. This why, in architecture, an inverted catenary curve is a useful shape for an arch. This form was applied by modernist architects and, in particular, by Antoni Gaudi.



Mathematical design of a catenary model and 2D catenary model for the design of Gaudi's Casa Vicens.

In an arch that acquires a catenary form, the tension that it supports at each point is spread between a vertical component which is the one that has to sustain the arch itself, and another pressure component, which is transferred by the arch towards the foundations without horizontal strains being created, except at the ends of the arch when it reaches the foundations. This property is distinctive and unique to this type of arch. It ensures that it doesn't require support at both sides to sustain it and it avoids any tendency to spread. In Roman chapels, thick walls were needed on either side of the doors and windows to prevent the rounded arches from cracking. Mediaeval architects never managed to design a perfect form for transferring the lateral strains, despite the Gothic pointed arches being much closer to the catenary form. They still required lateral support with strong buttress exteriors that absorbed and transferred the horizontal tensions to the foundations.

For the same length between suspended points, the parabola is less pointed than the catenary (it produces a bigger arrowhead as we said). If a load-bearing bridge adopts a parabola as its curve, the arrow (the lowest point of the bridge, in its centre) is slightly smaller than in a catenary. In a real suspended bridge we can consider that the cable mass is negligible with respect to the bridge and, therefore, a cable is considered to adopt a parabolic form.

In railway catenaries (overhead power lines), the upper and contact threads have the same order of mass; therefore, we could consider applying the above reasoning. However, in reality, there are very few suspension cables, and the upper one has the form of discontinuous lines.

The catenary Gateway Arch in Saint Louis, Missouri.

Computer-assisted design curves (CAD)

There are various types of curves that are drawn with computer-assisted design (CAD). Depending on their mathematical complexity they can achieve greater realism of the objects drawn. These 2D or 3D objects can then be printed on paper or even become part of an animation that can be presented as a digital film that simulates a 'virtual reality'.

The most elemental curves used in CAD are the so-called Bézier curves, in honour of its inventor, the engineer Pierre Bézier. To draw a particular curve through four points Bézier's cubic curve is used. If these four points are A(0,0), B(1,3), C(4,3) and D(6,0), then the curve (a third-degree polynomial function) written in an abridged form (in the form called 'vector') for values of u between 0 and 1 will be:

$$\text{BEZIER\_3}(u) = (1-u)^3 A + 3u(1-u)^2 B + 3u^2(1-u)C + u^3 D.$$

In a more developed equation, for the four points given the curve is as follows:

$$[x(u), y(u)] = (1-u)^3[0,0] + 3u(1-u)^2[1,3] + 3u^2(1-u)[4,3] + u^3[6,0]$$

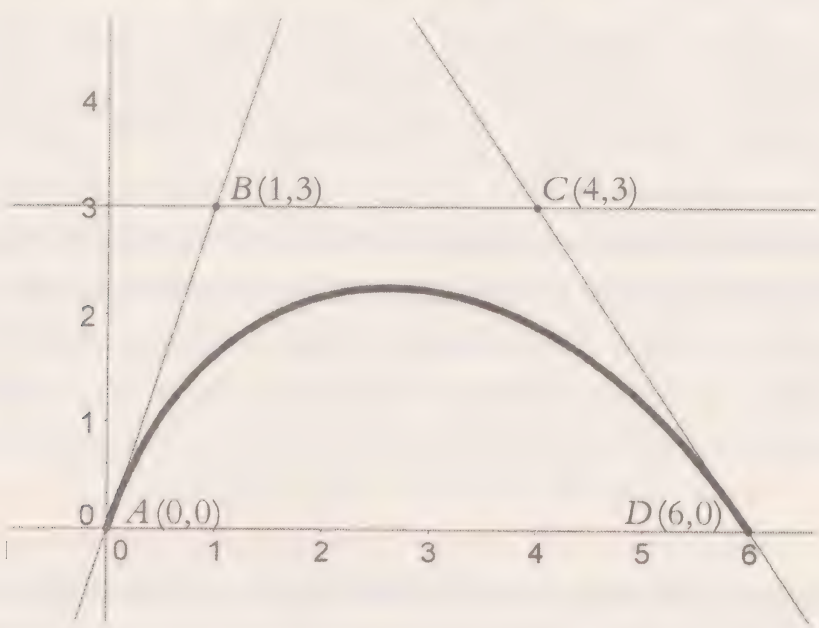
Separating the coordinates x and y two expressions would remain:

$$\begin{cases} x = (1-u)^3 \cdot 0 + 3u(1-u)^2 \cdot 1 + 3u^2(1-u) \cdot 4 + u^3 \cdot 6 \\ y = (1-u)^3 \cdot 0 + 3u(1-u)^2 \cdot 3 + 3u^2(1-u) \cdot 3 + u^3 \cdot 0 \end{cases}$$

After completing the indicated operations, these two equations generate a Bézier cubic curve in parametric coordinates:

$$\begin{cases} x = -3u^3 + 6u^2 + 3u \\ y = -9u^2 + 9u \end{cases}$$

It works as follows: the curve starts at point A and moves towards B; it then arrives at D coming from the direction of point C, as shown in the figure. Points A and D are crossing points for the curve that we want to draw; B and C are called control points. These points are used in such a way that the tangents of the curve drawn at A and D are the lines determined by AB and CD.



Bézier's cubic curve for points A (0,0), B (1,3), C (4,3) and D (6,0) with crossing points A and D, and control points B and C.

Normally, the curve will not pass through B or C; these points only provide directional information, or rather, directions of tangent lines at points A and D of the curve that we wish to draw. The distance between B and C indicates the length the curve should have when it 'moves' in the direction of C before heading to D in the same direction as the tangent line to the curve at D.

To improve the realism and smoothness of the object drawn, higher degree Bézier curves are used. To draw a particular curve with five points, for example, $A(1,1)$, $B(3,0)$, $C(4,5)$, $D(2,3)$ and $E(1,1)$, a quartic Bézier curve is used, in other words, a fourth-degree polynomial curve that can be written in an abridged fashion, for values of u between 0 and 1:

$$\text{BEZIER\_4}(u)=(1-u)^4A+4u(1-u)^3B+6u^2(1-u)^2C+4(1-u)u^3D+u^4\text{E}.$$

In a more developed vector form and for the five points given, the curve has this equation:

$$[x(u),y(u)]=(1-u)^4[1,1]+4u(1-u)^3[3,0]+6u^2(1-u)^2[4,5]+4u^3(1-u)[2,3]+u^4[1,1].$$

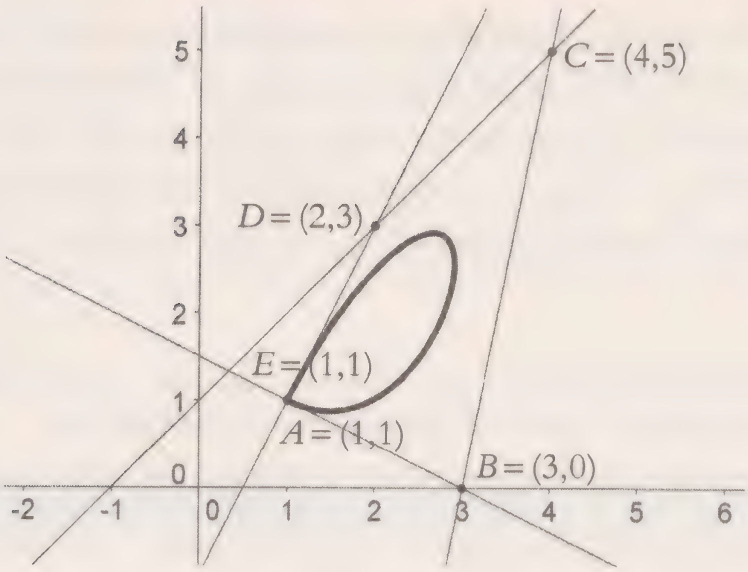
The curve starts at point A and moves towards B; then, C and D mark the form of the curve which approaches E coming from the direction of DE, as can be seen in the image opposite. Separating the coordinates x and y two expressions would remain:

$$\begin{cases} x=(1-u)^4\cdot 1+4u(1-u)^3\cdot 3+6u^2(1-u)^2\cdot 4+4(1-u)u^3\cdot 2+u^4\cdot 1 \\ y=(1-u)^4\cdot 1+4u(1-u)^3\cdot 0+6u^2(1-u)^2\cdot 5+4(1-u)u^3\cdot 3+u^4\cdot 1 \end{cases}$$

After completing the indicated operations, these two expressions generate the equation of a Bézier quadratic curve in parametric coordinates:

$$\begin{cases} x=6u^4-8u^3-6u^2+8u+1 \\ y=20u^4-52u^3+36u^2-4u+1 \end{cases}$$

It can be seen in the figure that the corresponding Bézier curve that is drawn lies within the polygon formed by the points that define it, which in the jargon of computer-assisted design tends to be called a 'convex capsule'.



Bézier's quadratic curve for points A *(1,1),* B *(3,0),* C *(4,5) and* D *(2,3) and* E *(1,1), with crossing points* A *and* E *and control points* B, C *and* D.

Other mathematical models exist for drawing more sophisticated curves with the computer, such as B-spline curves or NURB curves and surfaces. Splines are polynomial curves in sections with great continuity at the crossing points from one to another. The simplest example is the lineal spline in segments. It is simply a polygon line in 2D plane and in 3D space. Another examples is a cubic spline formed by third-degree curve arches.

The term spline means 'elastic slat', and makes reference to the slats that are used to create curves that described surfaces to design the hulls of ships and aeroplane fuselages. Subject to force, these elastic slats or splines assume a form that minimises their elastic energy. This property is adapted by third-degree mathematical splines. The geometric tool of splines is developed to solve the limitations of lack of local control that the Bézier curves had, their difficulty in imposing the continuity and imposition of the degree of curve according to the number of control points. A function $f(u)$ can approximate it using a polynomial p of grade 2 interpolating $f(u)$ in four abscissae: $u_0<u_1<u_2<u_3$. A result obtained in numerical analysis establishes the difference between p and f in the interval $[u_0,u_3]$ and can be expressed as:

$$p(u)-f(u)=\frac{f^{(4)}(v)}{4!}(u-u_0)(u-u_1)(u-u_2)(u-u_3),$$

where $v=v(u)$ it is in $[u_0,u_3]$.

Normally, the error is reduced when the differences between the abscissae decrease. However, an interpolator of a greater degree than the function $f(u)$ does not necessarily produce a better approximation of the same. For this reason, in many design programs it is common to use the polynomial functions in segments and at a reduced degree, such as cubic curves, which are often used due to their low degree and flexibility.

Curve that follows a set of points: interpolation

When we wish to discover a function or simple curve that adapts as well as possible to a particular set of points, in other words, which passes through all those points, the mathematical process is given the name interpolation. The simplest is polynomial interpolation whereby a polynomial curve is obtained that passes through all the points. The degree of the polynomial curve is lower than or equal to the number of points through which it has to pass.

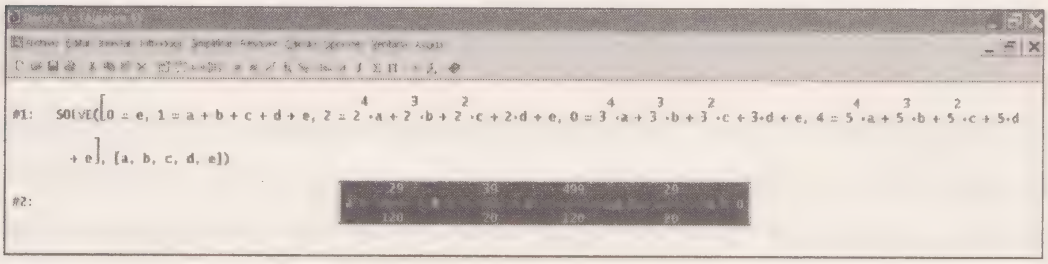
To resolve this issue there is a multitude of methods that have been proposed by mathematicians such as Newton or Lagrange. The simplest is the algebraic method, which implies the resolution of a system of linear equations (of first degree) in which the unknowns are the coefficients of the polynomial curve equation. For example, if we try to work out a curve that passes through the five point $A(0,0)$, $B(1,1)$, $C(2,2)$, $D(3,0)$ and $E(5,4)$, the simplest that meets this condition is a fourth-degree polynomial, or rather, a quadratic curve. Its equation has a maximum of five terms and looks as follows:

$$y = ax^4 + bx^3 + cx^2 + dx + e.$$

In the equation of the curve sought we don't know the values of the coefficients (the numbers that multiply the powers of the variable x), so we have given them the names a, b, c, d, e, which are the unknowns that need to be calculated. The unknown curve is $y = ax^4 + bx^3 + cx^2 + dx + e$, and it must pass through the five points $A(0,0)$, $B(1,1)$, $C(2,2)$, $D(3,0)$ and $E(5,4)$. A curve passes through a point when, substituting the x of the point (at point E is $x = 5$) in the equation of the curve, the result of the operations in the equation produces a y, which is the y of the point (that of point E is $y = 4$). The equation for each point is raised, or rather there are five equations like those expressed as follows, with five unknowns:

$$\left.\begin{array}{l} 0 = a0^4 + b0^3 + c0^2 + d0 + e \\ 1 = a1^4 + b1^3 + c1^2 + d1 + e \\ 2 = a2^4 + b2^3 + c2^2 + d2 + e \\ 0 = a3^4 + b3^3 + c3^2 + d3 + e \\ 4 = a5^4 + b5^3 + c5^2 + d5 + e \end{array}\right\}$$

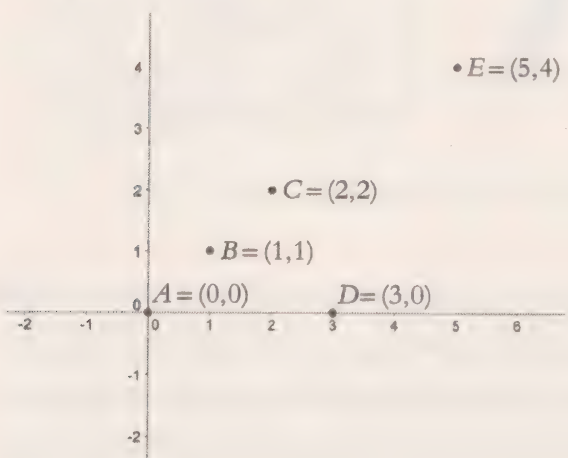
These equations are resolved with the symbolic calculation software Derive:

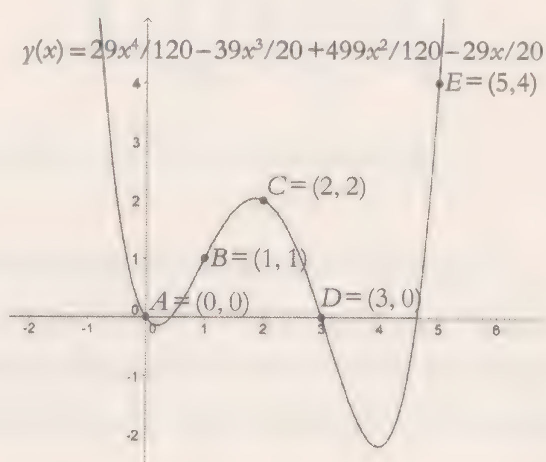


The solution is a curve with the equation:

$$y = \frac{29}{120}x^4 - \frac{39}{20}x^3 + \frac{499}{120}x^2 - \frac{29}{20}x.$$

In the figures that follow the points given can be seen together with the resulting curve that passes through all the points:





Curves in the design of typography

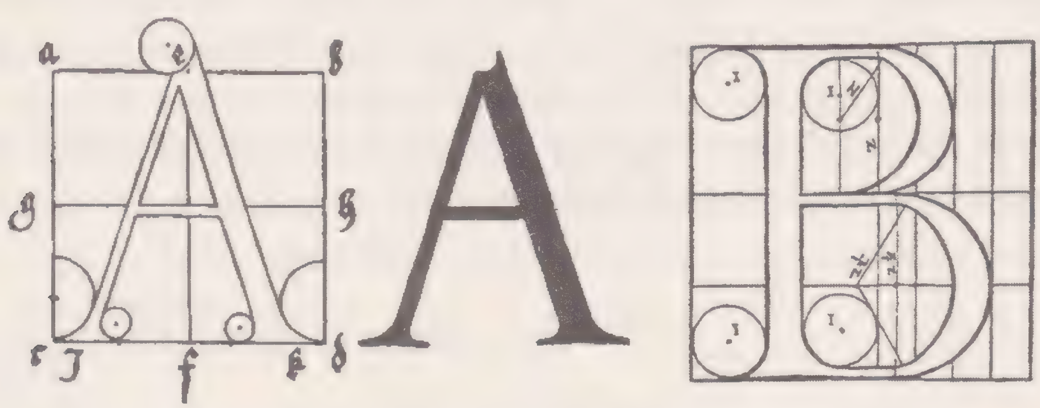
Typography belongs to fine art, where even the smallest details and elements are important, and it can influence the overall perception of the final creation, whether

it is an advertisement or a text printed with a certain type of font. From looking at the two types of font here, the importance of using curves in the design of typography can be seen.

Aa Bb Dd Kk Rr pPVv Qq Zz 1 3 8 % ? !
Aa Bb Dd Kk Rr pP Vv Qq Zz 1 3 8 % ? !

An example of a serif font (top) and sans serif.

The use of serif fonts is traced to the alphabets drawn in curves by the Renaissance artists, such as Dürer or the printer Francesco Torniello da Novara (1490-1589), who defined the characters of the ancient Latin alphabet inscriptions geometrically. Later on, he designed a grid of 18 × 18 points for each character, that served as coordinates for using his geometric fonts in printing types.



Upper case A by Albrecht Dürer (1528) and a B by Francesco Torniello (1517).

This is where this short tour through the house of curves ends. We have been able to review many aspects of our life where these marvellous and mysterious creatures are present. If the reader continues the journey through the lesser used passageways of this immense house, a world of endless marvels, surprises and adventures awaits you.

Bibliography.

BOLTIANSKI, V.G., *La envolvente*, Moscow, Mir Editorial, 1977.

HOGBEN, L., *Mathematics for the Million*, London, Merlin Press, 1989 (4th edition).

Lockwood, E.H. *A Book of Curves*, Cambridge, Cambridge University Press, 1961.

PEDOE, D., *Geometry and the Visual Arts*, New York, Dover Publications, 2000.

Revue du Palais de la Découverte Nº 45, *Courbes mathématiques*, Paris, 1995.

VASILEV N.B., GUTENMACHER V.L., *Straight Lines and Curves*, Moscow, Mir Editorial, 1980.

Index

Curious Curves

Ellipses, hyperbolae and other geometric wonders

A famous writer once distinguished between the "lines of the intellect" and the "curves of the emotions". Accepting this metaphor, readers should be warned that this book, despite dealing with mathematics, abounds with emotion on every page! Its elegant and sinuous heroines are known by suggestive names such as ellipse, hyperbola, spiral and conchoid, and their fine definition and seductive properties have turned the heads of mathematicians since the dawn of thought itself!